

REPRESENTATION WITHOUT REPRESENTATIONALISM

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Lucian Leahu

May 2012

© 2012 Lucian Leahu
ALL RIGHTS RESERVED

REPRESENTATION WITHOUT REPRESENTATIONALISM

Lucian Leahu, Ph.D.

Cornell University 2012

The idea that representations not only describe, but also help shape the world is being explored both empirically and theoretically in an increasing number of academic disciplines ranging from anthropology to quantum mechanics. Insights emerging from such research question representationalism: the belief that representations simply describe the represented. Ideas and arguments around the nature of representation are particularly relevant for computer science because computers are representational technologies: in order to be useful, they must represent relevant aspects of the world. In my PhD research, I have taken up the challenge of exploring the implications of these ideas for computational approaches.

Grounded in affective computing and ubiquitous computing, my research was guided by two core questions united by a focus on the opportunities and implications brought about by taking seriously the idea that representations also shape reality. The first examines how to derive computational representations differently by engaging this idea in technical practice. Two projects provide two case studies on different representations using physiological sensor data: one on basic visualizations of the data; the other, focuses on a more complex form of representation: training classifiers of emotion from the data using machine learning. Building on these projects, the second question examines how the shift in the way we understand representations changes the main practices of constructing interactive systems. To this end, I have designed, built, and

evaluated an interactive, mobile system leveraging sensor based statistical classification; the system is designed to help its users experience and understand their emotions.

BIOGRAPHICAL SKETCH

After graduating with a Bachelors of Engineering degree in Computer Science from Cluj Napoca Technical University, Romania in 2003, Lucian enrolled for graduate studies at Cornell University in the Computer Science department. There, he earned a Master of Science degree in 2008 and a Doctor of Philosophy degree in 2012. As for his future plans, Lucian will join the Swedish Institute of Computer Science as a postdoctoral researcher on an ERCIM (the European Research Consortium on Informatics and Mathematics) postdoctoral fellowship. He intends to continue to research interdisciplinary topics at the intersections of interaction design, computer science, philosophy of science, and anthropology.

Pentru bunicul Liviu si bunica Nica

ACKNOWLEDGEMENTS

I thank my PhD committee: my advisor, Phoebe Sengers, for her boundless support, incessant encouragement, sharp feedback, and kind words; Alex Taylor for suggesting a number of readings which turned out to reshape my thesis in important ways, for many inspiring conversations, and for his friendship; Geri Gay and Bob Constable for their unremitting enthusiasm for my work.

I feel lucky to have been encouraged over the years by many friends, near and far. I won't name you all, for fear that I will leave some out, but you know who you are. I'd be remiss if I didn't mention the van Loans, the Stones, the Carnies, and my CEmCom friends.

Throughout my time at Cornell I was very fortunate to have part of my research supported generously by the following organizations: Intel, Nokia, Microsoft Research Cambridge, Cornell's Graduate School, and Cornell's Computer Science Department.

Finally, my family has been behind me through thick and thin, in spite of never really feeling that they fully understand what exactly it is that I do.

A heartfelt thank you to you all!

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1 Introduction	1
1.1 Entangled Beginnings	3
1.2 Research Questions and Contribution	10
1.3 Overview	13
1.3.1 Chapter 2: The Socio-Technical Gap in Action	13
1.3.2 Chapter 3: There Has Never Been an Inherent Gap	14
1.3.3 Chapter 4: Representation without Representationalism	15
1.3.4 Chapter 5: Meeting Emotion Halfway: Designing, Building, and Evaluating an Affective Interactive System	15
1.3.5 Chapter 6: Conclusion	16
2 The Socio-Technical Gap in Action	17
2.1 Ubiquitous Computing	17
2.1.1 Context	19
2.1.2 Stuck on Either Side of the Gap	21
2.2 Affective Interaction	27
2.3 Gap Problems	31
3 There has never been an inherent Socio-Technical Gap	35
3.1 The Gap	35
3.2 Analyzing the Gap	38
3.3 Stepping Back	41
3.4 Understanding the World Differently	47
3.5 There Has Never Been an Inherent Gap	63
3.6 Performing HCI Differently	65
4 Representation without Representationalism	67
4.1 Preliminaries	69
4.2 Emotion as Emergent and Relational	72
4.3 Representing Emotion	75
4.3.1 Physiology and Emotion	76
4.4 Cutting Emotion Using Statistical Models from Physiological Data	96
4.4.1 What are Statistical Models of Emotion?	97
4.4.2 Goals of Study and Approach	99
4.4.3 The Study	102

4.4.4	Classification Experiments	125
4.4.5	Performing Representations Differently	136
4.4.6	Summing Up	141
5	Meeting Emotion Halfway: designing, building, and evaluating an af- fective interactive system	144
5.1	Preliminaries	145
5.2	Model Building	147
5.2.1	Sensors	148
5.2.2	Data Gathering	148
5.2.3	Model Training and Testing	152
5.3	Meeting Emotion Halfway: Designing Freaky	153
5.4	Hardware Implementation	159
5.5	Software Implementation	160
5.6	Studying Freaky in Action	160
5.6.1	Max and the cemetery	163
5.6.2	Uma and Tyler	165
5.6.3	Writing Anxieties	167
5.6.4	Intermezzo: Prefacing an Agential Realist Take on Realist Evaluation	168
5.6.5	Understanding Human-Machine Performances of Emotion	173
5.7	Conclusion	189
6	Conclusion	191
6.1	Limitations	198
6.2	A Critical Technical Practice	199
6.3	Looking Ahead	201
A	SVM Parameter Testing	203
A.1	Kernel Type	203
A.2	Time Window Size	204
A.3	Trade-Off C	204
A.4	Cost Factor J	205

LIST OF TABLES

1.1	Technical vs. socio-cultural approaches to human emotion. . . .	4
4.1	Comparison between personalized models and aggregate models for the ropes course data.	127
5.1	Overview of the users study.	162
A.1	Summary of results for kernel type parameter testing.	203
A.2	Summary of results for window size parameter testing.	204
A.3	Summary of results for trade-off parameter testing.	205
A.4	Summary of results for cost factor parameter testing.	205

LIST OF FIGURES

1.1	Arousal map — overlapped representations of the GSR data from several participants	6
4.1	Arousal map	79
4.2	Cornell campus mock aggregate map.	91
4.3	The ropes course.	105
5.1	Freaky's shell.	157
5.2	Interacting with Freaky.	157
5.3	One of our participants and Freaky.	161

CHAPTER 1

INTRODUCTION

Recurrent problems for interactive systems are rooted in specific differences between people and computers. Lucy Suchman's *Plans and Situated Actions*, a seminal text in HCI, takes such differences as a central concern (Suchman, 1987). Suchman grounds her analysis in observations of interactions between an 'intelligent' photocopier and its users. The copier, augmented with an expert help system, offered users step-by-step guidance on how to perform different kinds of copying jobs. To do this, it kept track of the user's actions, such as pressing buttons, and advised the user how to perform the next action by displaying instructions for her to follow. And yet, in spite of these technical efforts to ensure the smooth maneuvering of the copier, the encounter between novice users and the system often led to interactional breakdowns that could be summed up as users and system not being on the same page¹. Suchman's observations and analysis locate the source of these problems in asymmetries between users and the system in access to and understanding of the situation: "It was as if the machine was tracking the user's actions through a very small keyhole and then mapping what it saw back onto a prespecified template of possible interpretations" (Suchman, 2006, p. 11).

An anthropology of human-computer interactions, *Plans and Situated Actions* quickly became, and remains until today, a key reference for the entire field of HCI. Suchman's analysis spawned a subfield of HCI — Computer Supported Collaborative Work (CSCW) — to further investigate asymmetries between people and machines and propose solutions to overcome them. This is to say that

¹Suchman describes the difficulties as "two classes of communicative breakdown, the *false alarm* and the *garden path*" (Suchman, 1987, p. 163, original emphasis)

problems due to differences between people and computers, such as the ones highlighted in her book, have been a persistent feature in HCI's landscape. Difficulties associated with these differences came to be referred as the *socio-technical gap*: the divide between social requirements for system design and the technical means to support them. In other words, the gap refers to limitations that inhere in technical approaches to address the subtleties that characterize social life. Attesting to its weight for matters of interaction, the gap has been identified as the main intellectual challenge for CSCW and one of the fundamental problems for HCI, more broadly (Ackerman, 2000).

Suchman's quote above is suggestive of two main research directions that together aim to narrow the socio-technical gap. The first, a technical direction, is concerned with enlarging the keyhole, as it were: giving computers more information about their environments. For example, more powerful sensing equipment and more complete representations gives HCI systems more tools with which to ascertain users' actions, the system's context of use, and so on. The second direction complements the first and examines to how data are to be interpreted. For example, this direction investigates *the relationship between representations and entities in the world*. The challenge of exploring the nature of this relationship has been pursued primarily by social scientists. Suchman herself critiqued the idea dominating technical work at the time of her analysis that representations can and should mirror reality. Since then, the complex and nuanced relationship between representation and represented described by Suchman in the context of human action has been confirmed and extended to other human practices by further ethnographic observations. Yet, as I will argue in this thesis, more than two decades since its publication, the overwhelming majority of technical work still relies on the idea challenged by Suchman's

book: that representations should stand in one to one correspondence with reality. Why has this idea been so tenacious? If we take seriously the complexity of the relationship between representation and reality, how should technical practice change? How could we construct and use computational representations differently? Can a different take on technical representation alleviate the problems associated with the socio-technical gap? These are the main questions I take up in this dissertation.

1.1 Entangled Beginnings

To make more concrete the core issues associated with the interplay between the socio-technical gap and the nature of the relationship between representation and represented, which I henceforth refer to as the *problem of representation*, I begin with an example from my work which contains the seed that spawned my PhD research. It makes clear in a palpable way that the representation and the gap problems are linked in a constitutive way. While it suggests that considering them together might be valuable, it also foreshadows important roadblocks this brings up and must be addressed in order to move on.

The example comes from one of my first research projects in HCI. That research, carried out with my colleague Steve Schwenk, took as its starting point the use of real-time physiological signals, such as heart rate, to infer users' emotional state. This is typically achieved by developing algorithms that produce representations of emotions: technical objects that approximate the social object that constitutes the referent of representation, i.e., lived emotion. One example of efficient representations of emotion maps objective, observable signals to

Technical	Socio-cultural
Objective	Subjective
Well-defined	Situated
Generalized	Idiosyncratic
Discrete	Fluid
Formal	Mediated
Quantified	Interpreted
Represented	Experienced

Table 1.1: Technical vs. socio-cultural approaches to human emotion.

emotional states: for example, from biometric sensor data an emotional state is computed. Purportedly, such representations make it possible to continuously identify the emotion experienced by the person whose physiological measures are sensed. Interactive systems typically use these algorithmic representations of emotion to infer user's affective state and adjust the interaction with the user according to the inferred state, e.g., to avoid further frustrating the user. What I want to emphasize here is that these efficient representations of emotions are used in systems as equivalents of user emotions. In other words, these representations are de facto positioned in systems as technical objects that mirror social ones, i.e., emotions.

One problem identified in the realm of affective interactive systems pertains to the differences between these two types of objects, technical representations and socio-cultural understandings of emotion. An instance of the socio-technical gap, the problem, sometimes termed the **affective gap** (Boehner & Sengers, 2006), speaks of the dichotomies that arise when contrasting the technical and the socio-cultural in the arena of human emotions.

Table 1.1 (adapted from Sengers, 2006) gives a sense that the technical and socio-cultural approach emotion in asymmetric ways in almost every dimension

of interest and it points to difficulties interactive systems may encounter in use due to these mismatches.

Returning to our project, we aimed to explore the interplay between the problem of representation and the affective gap. Specifically, we wanted to investigate whether a more nuanced approach to the relationship between representations and reality may help us think differently about this gap. We began by conducting a participatory study. We recorded physiological data from our volunteers while they were walking on Cornell's campus. With the data we made simple representations: visualizations such as the one below. The data recorded consisted of galvanic skin response (GSR) and the participants' location. Physiological research understands GSR, which measures the skin's conductivity, to be a measure of physiological arousal, which plays a central role in many theories of affect. The map visualized the GSR data with the users' trajectories (figure 1.1): the height of the wall corresponds to the GSR value recorded at that location.

In the discussions with the participants that followed the data recording, our aim was to observe how they made sense of their experiences and actions and how the maps factored into these activities. The participants gave rich accounts of what had or might have happened to trigger the physiological events depicted in the maps, e.g., peaks in the GSR values. Importantly, the maps were weaved into their accounts together with other resources, experiences, and memories which allowed people to generate new understandings and meanings of their experiences. In chapter 4, I will present at length the themes and insights that emerged from this study; here, I highlight a few related points that are particularly relevant in framing the focus of this dissertation.

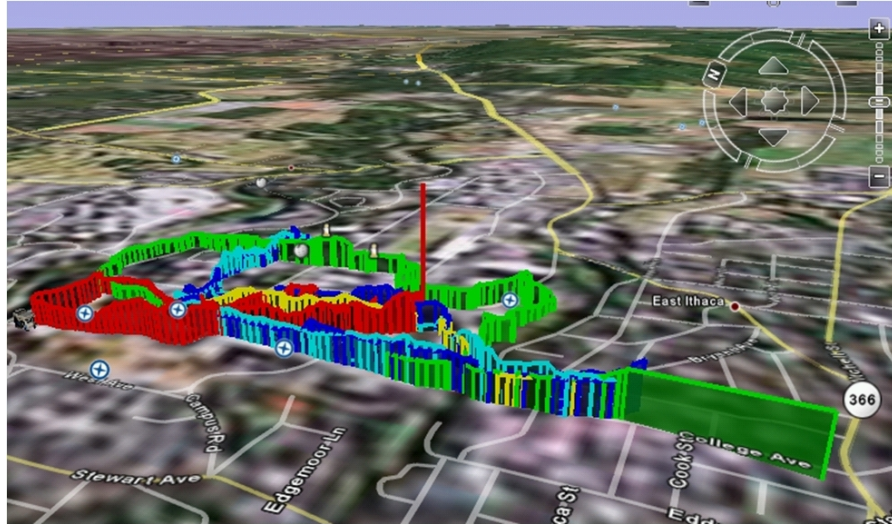


Figure 1.1: Arousal map — overlapped representations of the GSR data from several participants

The most vivid impression the study made on us was that when the meaning of representations such as the maps is not fixed a priori, wonderful things start to happen. As people contextualized and recontextualized these representations in their narrations, the representations became in an important sense alive: generative and dynamic, as evidenced by the breadth and depth of interpretations. That is, the maps participated in people's accounts beyond the straightforward referential relation that would be typically expected of such representations. Importantly, people's accounts seamlessly weaved the representation into their own accounts, emotional and otherwise. In other words, no issues related to the affective gap surfaced in people's interactions with the maps. For us these observations suggested that in dealing with problems associated with the affective gap, reconsidering the nature of the relationship between the representation and represented might offer solutions.

This study also included a design component in which we aimed to con-

ceptualize interactive systems embodying the sense of dynamism that emerged from our discussions with the participants. study and to engender new possibilities for affective interactive systems. The artifacts and systems we envisioned demonstrated a different design space afforded by a more nuanced relationship between representations and represented that moves beyond positioning representations as mirroring lived emotions. Rather, we used representations to provoke and to encourage users to engage critically with them, e.g., by contesting their meaning.

As I presented this work to different audiences, I started to become aware of serious differences in how these issues were understood from different disciplinary backgrounds. The paper describing the project was published at DIS, a design oriented HCI conference typically attended by researchers from a medley of backgrounds brought together by a sensibility towards matters of design. Interdisciplinary audiences such as the one at DIS showed excitement and interest in this work. Interestingly, the research was received differently by more technical audiences, such as artificial intelligence researchers, accustomed exclusively to using representations as correspondents to real-world entities. While some members of the audience appreciated this new sense of freedom around representation, most of them were decidedly uncomfortable. I was quickly engulfed by an avalanche of related questions that would be best summed up as pertaining to ground truth: “what about the true meaning of the physiological data?” “These interpretations are nice stories that people tell, but how do we know these people aren’t simply deluding themselves?” “Some of the interpretations appear to be contradictory. Clearly, only some of them may be truly true, the others are simply false and it’s our job as scientists to clarify things for users and to do so we must go back the emotion experienced at the

time the data was recorded.”

I found comments and questions such as these disconcerting. Talking about the correct emotion seemed to move the discussion away from what was important for our participants. My sense was that questions of ground truth (the emotion experienced at the time of the recording), while important, are secondary to the connections the participants made and the meanings they arrived at when coming up with and sorting between different interpretations. Yet, from the traditional technical perspective embraced by my audience, it felt as if the labels (i.e., true emotion) corresponding to each data point were simply erased or hidden from view in my study.

Another unsettling aspect for me was that the general agreement among my more technical audience appeared to be that while this research could be useful for matters of human-computer interaction, computational and more specifically artificial intelligence research such as emotion recognition through pattern matching from physiological data need not worry about matters of interpretation. From this perspective, the interpretational approach I was suggesting could just be added to the interactional layer of the system, i.e., on top of the technical level, if such an approach proves to be useful for users. In other words, the overall sentiment seemed to be that computational efforts are the ones handling the object of investigation (true emotion), so they need not worry about broader social aspects surrounding the system such as user interpretation; these were issues that could be left for the interface.

As a computer scientist and given my research background in artificial intelligence (AI), I knew exactly where my AI colleagues’ questions were coming from. My study was suggesting that a way out of recurrent difficulties asso-

ciated with the affective gap – in short, differences between representations of emotions and lived emotion – was to approach differently the relationship between representations and the world. My study was going against what in computer science is a matter of common sense: the assumption that technical representations should stand in direct correspondence to the entities they represent; therefore, representations of emotions should correctly mirror users' actual emotions, in contrast to the fuzzier, more entangled sorts of relationships that seemed to be suggested by users' entangled accounts. From their perspective – which I had shared for many years – it is very hard, if not impossible, to imagine what this might mean. Wouldn't letting go of the correspondence assumption lead to utter relativism? Would anything go?

My colleagues' questions made me realize that in order to move forward with my work on different takes on representation and the socio-technical gap I would first have to answer some important questions, although not quite the ones my audience was asking me. In a nutshell, an important element of the study was that representations not only describe aspects of the world, but they also shape them. In other words, I was convinced that the nature of the relationship between technical representations and the world must be revisited. While my intuition was that a rigorous technical practice is still possible if we let go of the correspondence assumption, I did not know what such a technical practice might look like, and neither did my audience. Moreover, these considerations also had philosophical undertones, as reflected in the kinds of questions that my interlocutors asked. For example, if representation is no longer to be considered as mirroring entities in the world, then what is the objective referent of representation? What about scientific truth (e.g., laws of nature) if we loosen the direct correspondence between representation and represented, does that mean

we no longer have access to truth?

What I was missing at the time was a coherent framework in which these and other related issues could be explained and which would provide a conceptual foundation for rethinking and performing scientific and technical work without fear of relativism and anything goes.’ Without this framework, my work would simply be interpreted – within the existing framework which includes and supports the commonsense view of representation – as a playful, arty engagement with representation in which one pretends to have forgotten, or does not need to figure out, the precise, correct referent. Within the classic framework for thinking about representation, then, insights such as the ones that emerged from my study may appear inconsequential to technical practice. To put it differently, within the classic framework palpable differences that may unlock recurrent problems, such as generative and vibrant capacities of representations, are likely to be explained away. It became clear that a different framework was needed that would allow me to account for these differences and insights differently.

1.2 Research Questions and Contribution

The project presented in the previous section suggests an entangled relationship between a variety of issues that surface in HCI practice. These include practical limitations of interactive systems, such as the ones described by the socio-technical gap; core technical concerns, e.g., how to construct more useful computational representations; conceptual issues pertaining to the nature of the relationship between representation and represented; and, epistemological

ones: questions of validity, objectivity, and truth. My first aim in this dissertation has been to clarify the interplay between these matters. A better understanding of how they relate to one another would allow us to act differently – which constitutes a second goal for my work – for example, in proposing fundamentally different solutions to the problems associated with the socio-technical gap or in constructing more useful technical representations. Concretely, my research addresses the following questions:

- **How is the socio-technical gap related to the problem of representation?**

I will show that the disjuncture between the social and the technical is entangled with a web of metaphysical assumptions about what it means to know the world. Sometimes referred to as the tripartite separation (e.g., Barad, 2007; Verran, 2001), this web of assumptions postulates the a priori separation of knower, knowledge, and the world. A core aspect that supports the separation is *representationalism*: the belief that representations can and should mirror entities in the world and that the represented exists independently of practices of representation. The two issues, then, are entangled with one another and with a web of metaphysical assumptions.

- **How may we overcome the gap?** As I suggested earlier, reconsidering the relationship between representation and the world holds the key for reframing the problems associated with the socio-technical gap. But in itself this move is not enough, as representationalism is at the core of the metaphysical framework that underlies the majority of work in HCI. Therefore, important philosophical questions must be answered and an entirely different metaphysical framework is needed: one that does away with representationalism, and therefore allows a different take on the nature of representations.

- **How do we conceptualize the world differently?** Representationalism has been problematized by empirical as well as theoretical results in a number of disciplines, including anthropology, feminist theory, philosophy of science, and quantum mechanics. Performativity has been put forward as a potential alternative to representationalism. An important distinction between the two approaches pertains to how separations such as the ones between representation and represented or social and technical are conceptualized. In contrast to representationalism which relies on the separations being inherent in the world, performativity sees separations as being achieved through specific practices. Therefore, they may be performed differently. In the context of the socio-technical gap this would mean that our current practices enact the two categories as separate and distinct; therefore, changing our practices may allow the two to be co-articulated. I will draw on a philosophical framework that is based on performativity and offers a solid foundation for rigorous scientific and technical work: Karen Barad's agential realism (Barad, 2007). I adapt her framework to show how it can be applied in HCI; the adaptation is a key part of my contribution.

Having clarified important conceptual issues, in the second half of the dissertation I outline how HCI practices begin to change when looked at through this philosophical framework. Concretely, I will address the following questions:

- What are the consequences for technical practice of adopting a performative approach? How do we perform technical representations if they may be no longer assumed to mirror the world?
- What is the nature of the socio-technical gap within the new framework?

- How do we design, build, and evaluate systems differently in light of the new framework?

1.3 Overview

An overview of each chapter demonstrates where my research has led and how the chapters fit together.

1.3.1 Chapter 2: The Socio-Technical Gap in Action

The socio-technical gap has become an increasingly common, sometimes invisible, lens through which the HCI research landscape is implicitly viewed and approached: research questions pertaining to the gap drive the field's research agendas, the gap shapes the kinds of solutions we explore, and evaluation of new methods and systems is performed relative to the gap (e.g., have we narrowed the gap?). In this chapter, I highlight the gap's pervasiveness in structuring both discourse and technical practice. To do this, I provide a conceptual overview of the fields of ubiquitous computing and affective computing, the target areas for the studies that make up this dissertation. This overview focuses on the interplay between the gap and strategies to address issues associated with the gap. I show that efforts to bridge the gap are stuck on either side of the divide and that the gap is fundamentally entangled with the problem of representation.

1.3.2 Chapter 3: There Has Never Been an Inherent Gap

The nature of the socio-technical gap constitutes the focus of this chapter. HCI understands the gap as real and inherent in the world. I open with an overview of the way this came to be and how HCI as a discipline has theorized and positioned the categories of social and technical. Following that, I examine the gap in the context of the broader philosophical assumptions that it rests upon. This will allow us to understand the problem more deeply – why we feel stuck on either side of the divide – and its origins. My analysis shows that the philosophical foundation the gap rests upon creates the conditions for the deadlock described by the gap. Rather than being inherent in the world, then, the gap is an artifact made possible by the very ways we have come to see the world.

In the second half of the chapter, I show that a fundamentally different philosophical framework is needed to transcend the gap. To this end, I introduce an alternative framework – Karen Barad’s agential realism (Barad 2007) – which sees separations, such as the one between social and technical, as performed through specific practices, as opposed to an inherent property of the world. While dissolving the view of the gap as inherent, this realization does not dissolve the differences between the social and the technical. Recognizing the co-articulation of the social and the technical, in other words, does not mean that there are no differences. Rather, it situates those differences as continuously enacted, rather than fixed, and therefore subject to different enactments. The remaining chapters explore concrete technical and design strategies to create opportunities for the social and the technical to be co-articulated.

1.3.3 Chapter 4: Representation without Representationalism

Two studies form the core of this chapter. They provide two different cases of using physiological measures in the context of representations of emotions. The first, presented briefly in this introduction, visualizes the data and presents it to the participants in a map form. The second uses the data to build machine learning models that predict the user's emotion from physiological measures. I use these projects to demonstrate concretely what it means to perform and use representations without relying on representationalism – the assumption that representations can and should reflect entities in the world. This is not a simple task, given that technical representations are intimately associated with, if not founded on, the idea of representationalism. Yet it is important that we do so if we are serious about transcending recurrent difficulties, such as those associated with the socio-technical gap. The aim of this chapter is to show what it might mean to approach representation within the alternative framework introduced in the previous chapter, agential realism. This chapter, then, is about understanding, performing, and using representations differently.

1.3.4 Chapter 5: Meeting Emotion Halfway: Designing, Building, and Evaluating an Affective Interactive System

This chapter builds on the previous ones and describes the design, construction, and user study of an interactive system performed within the alternative conceptual framework proposed in chapter 3. The system I built exploits the technically advanced representations of emotions that were the object and vehicle of investigation in chapter 4: statistical models that classify physiological

data from sensors into emotional categories. The system – named Freaky – is an interactive, mobile system engaging sensor-based statistical classification; it is designed to help its users experience and understand their emotions. Freaky offers a demonstration of what building, designing, and evaluating from this alternative perspective entails how all these steps are performed differently, or not, vis a vis traditional approaches. Moreover, performing all these different steps provided another opportunity to engage further with the issues that are central to this thesis: the problem of representation, the perceived distance between the social and the technical and, ultimately, adjustments to our practices that may engender more compelling human computer interactions.

1.3.5 Chapter 6: Conclusion

I conclude with an overview of the main themes and contributions of this research. I detail why the reframing proposed here might be attractive for technical and design practices in HCI and discuss limitations of taking such a reframing.

CHAPTER 2

THE SOCIO-TECHNICAL GAP IN ACTION

The disjuncture between social requirements for technologies and the technical means to support them is a key concern in the CSCW community. This disjuncture came to be named the socio-technical gap. The gap, however, extends beyond collaborative systems. Indeed, it has become an increasingly common, sometimes invisible, lens through which the HCI research landscape is viewed and approached: questions pertaining to the gap drive much of the field's research agendas, strategies to answer such questions are shaped by the gap, and evaluation of new methods and systems is performed relative to the gap (i.e., have we narrowed the gap?). This chapter highlights how HCI typically renders the gap. It does so by giving a sense of the gap in action: the two sides of the gap, approaches on each side, and the impasses these give rise to. I trace these by grounding the discussion in two subareas of HCI that set the scene for the studies that make up this dissertation: ubiquitous computing, a broad research avenue in HCI that addresses computation beyond the desktop, and affective computing, unquestionably a socio-technical endeavor and therefore an apt area for making visible the gap and issues around it.

2.1 Ubiquitous Computing

In the past two decades, a major research area within HCI has been exploring new opportunities for interaction by integrating computational devices in everyday human settings. This area is typically referred to as ubiquitous computing (Weiser, 1991), although it is also known as pervasive computing (Ark &

Selker, 1999), context-aware computing (Dey, 2001), proactive computing (Tenenhouse, 2000), ambient intelligence (e.g., Aarts, 2004), embodied interaction (Dourish, 2001), etc. The different names suggest different aims and approaches to embedding computation in human environments. In spite of differences, these facets of ubicomp come together around the aim to extend interaction beyond the desktop setting and beyond the one user–one computer paradigm. The continuous decrease in cost, size, and power requirements for sensing devices and computational platforms has triggered an enormous interest in technologies that could leverage these devices, advancing interaction into virtually every space populated by people (e.g., kitchens, museums, subways) and beyond (e.g., the seabed, or the planet Mars). As people and devices are increasingly mobile, ubicomp envisions the world itself as the ultimate interface to computation (Weiser, 1991).

Liberated from the confines of the office or the desktop, computational devices now find themselves inhabiting heterogeneous environments and interacting with diverse user populations. Indeed, ubicomp systems have been designed for settings ranging from hospitals (Bardram et al., 2012) to playgrounds (Lo et al., 2007) and for users groups including firefighters (Ramirez & Dyrks, 2010), commuters (Bassoli et al., 2007), and children diagnosed with autism (Kientz et al., 2005). Moreover, devices such as smart phones accompany users pretty much everywhere. This explosion in the range of settings introduces new technical and design challenges for ubicomp systems. Perhaps the most important challenge for technologies in this arena has been to adapt to the dynamic nature of human environments. Adaptation in this context has typically meant making sense of the data available (e.g., from sensors) so that interaction may be more sensitive and responsive to its setting. In other words, the context of

interaction becomes a central aspect for ubicomp systems.

Context is by no means a novel concern in HCI. Indeed, the CSCW community has made the situated nature of human activity a central, if not the main, research topic. Situated action — as understood by social scientists within and around CSCW — points to the importance of context in making sense of human action: meaningful actions cannot be understood separate from the context in which they emerge (Suchman, 1987). I continue by dovetailing the technical position on handling context in system design with social understandings of context.

2.1.1 Context

The most common approach to matters of context in ubicomp has been to define what context is and to represent it. Concretely for a particular application, this might involve the following steps: identifying aspects of the environment that are relevant for the system to be designed, e.g., users' location and current activity; endowing the system with sensors that would offer access to information pertinent to those aspects, e.g., GPS technologies, proximity sensors, cameras; internally representing a mapping from sensor input to context dimensions, e.g., GPS location in a particular range means the user is at home; and, encoding the appropriate reaction or response of the system for that particular context, e.g., if home and having dinner, set phone profile to silent'.

Building on this basic idea of representing context more sophisticated approaches have been proposed, such as: sensor fusion, combining data from multiple sensors so that the result is more accurate and reliable (Dey et al., 1999);

context tool kits, essentially encapsulating sensing which may then be accessed through an API by higher system levels (Salber et al., 1999); employing predictive models for attributes of interest, e.g., using machine learning to extract patterns in the data and use them to recognize categories of interest from sensor data or system input (Tapia et al., 2004); using hierarchical decomposition of human activity into smaller, more basic activity so that user action may be inferred from observing these smaller actions (Li & Landay, 2008), etc. The picture that emerges is one in which the physical, observable aspects of the environment are used to give applications information pertaining to the use environment.

Approaches to context such as these have shown mixed results. I refer to these efforts as representationalist, as they rely heavily on the one-to-one correspondence between representations and the entities represented. By and large, systems embodying such strategies are able to function reasonably well in settings that are characterized by little uncertainty, for example, factory automation lines or laboratory settings, i.e., in settings in which potential sources of ambiguity or uncertainty have been minimized. While such results can be seen as encouraging, the persistent lack of systems demonstrating such approaches in everyday (unconstrained) casts a shadow over these methods (e.g., Rogers, 2006; Leahu et al., 2008b).

Part of the problem lies in what is taken to be context: what is relevant for the current action. On the technical side, representing context has been limited by what can be sensed in the environment. This has meant an emphasis on physical, observable aspects. Crucially for the topic in this chapter, social scientists link meaning in human environments with more than just physical or material aspects. For example, Suchman shows that for action to be intelligible

it must be situated in collective social and cultural practice (Suchman, 1987), while Chalmers posits temporal and personal dimensions to context: our past experiences influences what and how we do in the present (Chalmers, 2004). We begin to see what Dourish calls a mismatch between social and technical aspects in ubicomp (Dourish, 2004). Thus a gulf emerges between the nuances and subtlety of context as described by social scientists and the technical possibilities to capture it in/for interaction.

2.1.2 Stuck on Either Side of the Gap

As I alluded to earlier, both technical and social perspectives are in agreement on the lack of palpable results of overwhelmingly representationalist technical approaches to context in real, everyday settings (e.g., Abowd & Mynatt, 2000; Intille et al., 2004; Rogers, 2006; Leahu et al., 2008b). The two sides, however, disagree about what this lack of progress means and, consequently, on how to remedy this situation.

The technical community is increasingly aware that representations of context fail to capture context in its fullness. Part of the problem is thought to stem from technology's limited access to other aspects of context that do not leave measurable traces, for example cultural differences and sensibilities. At the same time, it is hoped that representations of context will become good enough approximations to get the job done. Part of the argument is that success in controlled environments demonstrates that such representations can be made to work in real environments; if they have yet to make it in real settings, it is only because ubicomp technologies are not quite there. Indeed, in a highly cited

overview of ubicomp research directions, Abowd and Mynatt note: “[w]ithout good representations for context, application developers are left to develop ad hoc and limited schemes for storing and manipulating this key information. The evolution of more sophisticated representations will enable a wider range of capabilities” (Abowd & Mynatt, 2000, p. 37). Examples from the ubicomp literature of target areas for improvement include:

- improving sensing capabilities in real settings; currently, sensors that work well in limited settings prove faulty and brittle in the messiness of everyday human environments.
- coming up with better ways to process sensor data, e.g., different statistical features may be relevant for unconstrained settings.
- developing representations that are culturally specific, e.g., studying ethnic differences in physiological response to emotional stimuli.

The picture that emerges on the technical side of the gap, then, is that a brighter future for ubicomp systems lies just around the corner, by simply continuing along the same general direction of formalizing context and creating good enough, workable approximations of context.

The social bank of the gap is less homogeneous in its stance towards the gap. This has to do with the different analytical traditions to which social scientists adhere. Dourish describes three kinds of conceptual orientations held by social scientists relative to their conceptual orientation: positivist, phenomenological, and critical. Positivist social scientists “seek objective, independent descriptions of social phenomena, abstracting from the details of particular occasions or settings, often in favor of broad statistical trends and idealized models.” (Dourish,

2004, p. 20). This kind of prescriptive knowledge fits well with the representationalist approach to context. The social-technical gap is productive for this kind of social science, in the sense that it continuously generates work, for example, to capture prescriptive aspects of human environments and practices to guide technical interventions; to document the mismatches between the social and technical, which are to be bridged through technical work.

From the phenomenological and critical perspectives, representing context is fraught with two kinds of problems. First, the transition from controlled to everyday settings is viewed as more problematic than is assumed by the technical camp: similar efforts to adapt representationist approaches to everyday settings have plagued the field of Artificial Intelligence since the 70s (Rogers, 2006; Leahu et al., 2008b); thus if it continues along current representationist strategies, ubicomp's bright future can be thought to lie perpetually just beyond the horizon line, rather than around the corner. Second and more fundamentally, the relationship between action and context that underlies the representationist approach is simply seen as misguided (Dourish, 2004). In the following, I focus on this latter concern as a means to uncover conceptual differences across the social-technical divide.

In *"What we talk about when we talk about context,"* Dourish takes a closer look at the mismatch between the social and the technical in ubicomp. He examines the relationship between context and action implied by representationist approaches:

- "context is a form of information": it can be extracted from a situation and encoded in the system
- "context is delineable": the context of use for a particular application can

be known and thus specified in advance

- “context is stable”: the context may change from one system to another, however for a specific system it does not vary between different usage instances. In other words the context specifications for a system are static.
- context and activity are separable”: context is seen as an outside factor to activity and the boundary is assumed to be clear.

In contrast, drawing on the phenomenological tradition Dourish understands this relationship differently:

- context is a relational property between activities and the surrounding aspects. This means that an attribute of the environment is not simply part of context (or not), rather it emerges as a relevant attribute (or not) in and through action.
- the boundary between action and context is fluid (or dynamic), rather than being delineable.
- context is particular to each instance or repetition of action, rather than being stable.
- context and action mutually constitute each other, rather than existing independently and thus being separable. In other words, context cannot be said to exist a priori, as it arises with the activity.

Dourish concludes based on these deep conceptual differences that “these two positions are incompatible” (Dourish, 2004, p.21). He proposes an alternative for ubicomp based on phenomenological understandings of human action and intelligibility. In his alternative, the goal of ubicomp systems becomes

supporting “the process by which context is continually manifested, defined, negotiated and shared” (p. 26), rather than modeling an external, observer-independent context. This alternative cannot rely on encoding meaning in the system, as meaning is inextricably linked to practice and therefore it cannot exist by itself; rather, it could facilitate “the evolution of meaning within communities of practice” (p. 26). Thus context is recast from a representationist to an interactional problem.

An example of a system which can be thought to embody this alternative approach to context is the Tableau Machine (Pousman et al., 2008). Designed to play the role of a non-human actor in the home, the system aims to provoke long-term reflection on daily activities and the social atmosphere in the home. To do so, the Tableau Machine uses sensed data to compute abstract measures of social activity in the home. Overhead cameras sense motion, which forms the basis for calculating measures such as social density as a way of continuously characterizing people’s activity in the home. This information drives a graphic engine which generates dynamic visuals projected on a large LCD mounted in a common space in the home, such as the living room. The system artfully handles the fluid boundary between activity and context: the system’s view of social activity becomes part of the physical and social context as it is displayed on the wall; at the same time, contextual information communicated through the display becomes the center of attention and an active part of the activity itself when people interpret it. Thus, the system can be thought to embody a different, more dynamic relationship between context and activity. This technical strategy avoids formally identifying aspects of the environment: indeed, the Tableau Machine characterizes dwellers’ activity, in contrast to the representationist strategy of automatically ‘recognizing’ activity from sensed data.

Systems such as the Tableau Machine demonstrate that non representationist approaches to context may be useful in everyday environments. Yet, from the technical bank of the gap they are often considered closer to interactive art than to ubicomp most likely because such alternatives have been slow to enter the domain of ‘serious’ applications, although notable exceptions do exist (e.g., a system for firefighters Ramirez & Dyrks, 2010). I return to the question of the limited impact of this alternative in chapter 3; here I want to note that Dourish’s conceptual alternative casts representationist approaches to context as forbidden: because context is not a thing that can be captured, representations of human environments are misleading and therefore systems relying on them are headed in the wrong direction, as it were. Relative to the socio-technical gap, this move may appear to firmly position the alternative on the social side of the divide and thus to solidify the separation of the technical and the social.

So far in this chapter, I have laid out conceptual, social, and technical considerations at the core of the field of ubicomp. I have shown social understandings of everyday human environments are at odds with the dominant technical strategies to build interactive technologies for those environments. In the next section, I turn to affect and the way it has been handled in interaction design. As it will become clear shortly, this domain exhibits a similar split between the social and the technical, similar conceptual clashes between the two, as well as related representationist and non-representationist approaches to those described in the context of ubiquitous computing.

2.2 Affective Interaction

Affect officially entered the realm of computing in the 1990s. Motivated in part by discoveries in neuroscience that too little emotion can impair problem solving abilities in humans, computer scientist Rosalind Picard has argued that providing computers with emotional abilities will make computational technologies more efficient and better at relating to human users. Her book, *"Affective Computing,"* puts forward a vision for "computers to recognize, express, and "have" emotions, as part of efforts to make them more intelligent, friendly, and capable" (Picard, 1997, p. 85). Emotion is thus positioned as another dimension of human environments which computers should tap into to participate more fully in them.

Drawing primarily on findings from physiology, psychology, and neuroscience, Picard and others have approached emotion from an engineering perspective. For example, laboratory studies of emotion in physiology have examined how the human body reacts during emotional stimuli. Work in this area has identified a number of measurable properties — like heart rate; galvanic skin response, i.e., skin conductivity; respiration volume; brain waves — that show variation with emotion. Picard posits that physiological measures can be recorded continuously via sensors and thus can be approached by computers as "signals that carry affective information" (Picard, 1997, p. 142). In this way, standard methods from signal processing can be applied to biometric sensor data. Moreover, pattern recognition and machine learning approaches may be leveraged to find emotion-specific patterns in the data, which can then be used by interactive systems to recognize the user's emotion.

An example of technology demonstrating these approaches is Wayang, an affect-aware tutor that teaches math and prepares students for standardized tests (Woolf et al., 2009). The system manipulates a variety of sensor data to infer students' interest and adapt the interaction accordingly. More specifically, facial muscle movement tracked from a video camera stream allows the system to infer facial affect, e.g., smiling, frowning, etc.; a chair fitted with a pressure seat gives the computer access to the student's posture; a sensor on the mouse detects changes in applied pressure which may correlate to student frustration or excitement; a wireless bracelet measures the galvanic skin response, an indicator of emotional arousal. A complex algorithmic architecture aggregates the patterns from all four sensors and predicts student interest. This information drives an animated affective agent's behavior who guides, encourages, and advises the students.

The affective tutor exemplifies the dominant technical approach to affective interaction: emotion is explicitly inferred and represented in the system. Boehner et al have argued that this approach casts emotion as another type of information and affective interaction as an information transmission problem (Boehner et al., 2007). Giving emotion a life in the digital realm has been achieved by fitting emotion to existing technical strategies: affective interaction is made equivalent to signal transmission through a noisy medium. In this model, the person experiencing emotion sends it to the receiver — a computer or another person — encoded in observable or objective, measurable signals; the receiver decodes these signals to recover the sender's emotion. In other words, emotion is seen as a biological fact and people as potentially noisy carriers — people's self reported emotions may be contaminated by social conventions which might dictate hiding one's true, biological emotional response, e.g.,

expressing affection in a professional setting may not be appropriate.

Boehner and her colleagues' account of affective computing methods and strategies shows that viewing emotion as information means treating it as an individual, internal, and delineable phenomenon. They contrast this view to socio-cultural perspectives drawn from anthropological and historical accounts of emotion. These perspectives locate emotion in shared, culturally embedded, social practices. Culture and social practices, then, are not simply a mask positioned on top of true, biological emotion, but the very place where what it means to have an emotion is continuously reworked. It follows that, in this view, emotion emerges as a relational construct in and through social interaction, rather than occurring independent of it, internal to the individual's body and mind.

Returning to technical approaches to emotion, these socio-cultural understandings problematize formal representations of emotion which are the basis of affective computing and its view of emotion as information. Specifically, seeing emotion as "culturally grounded, dynamically experienced, and to some degree constructed in action and interaction" (Boehner et al., 2007, p. 276) emphasizes the relational, subjective, and necessarily ambiguous nature of emotion, in contrast to the scientific as well as technical view of it as internal, individual, delineable, and objective. From a socio-cultural perspective, then, technical approaches leveraging representations of emotion construct emotion to fit scientific and technical approaches. In other words, a gap separates technical approaches to emotion from social, i.e., lived, phenomena.

Boehner et al. point out that representationist approaches shift the focus from the work emotion does in interaction to the identification of the experi-

enced emotion. This shift is a consequence of the signal-transmission approach to emotion: for the information to be transmitted it must first be identified. Furthermore, evaluating the success of representationist approaches means verifying that the person's true, experienced emotion was correctly recognized by the technology. For example, to gauge progress on the technical side, technical papers report improvements in recognition by contrasting the percent of correctly identified emotion of a new vis a vis existing techniques. In other words, the aim on the technical side is to close the emotion recognition gap by providing more efficacious representations of emotions.

From the social side, a different picture emerges. In everyday interaction knowing one's true emotion is secondary to achieving mutual intelligibility. By focusing on the former, technical approaches to emotion effectively move the discussion from what is important in the (social) world to that which is central in the scientific and technical world. The shift in focus seems to reinforce the split between the two, rather than bringing them closer.

Seeing representationist strategies as misguided, Boehner et al and others (e.g., Höök et al., 2008; Sengers et al., 2008; Gaver et al., 2007; Sengers & Gaver, 2006; Boehner et al., 2008) put forth an alternative approach to affective interaction that puts at its core a socio-cultural view of emotion. This alternative recasts emotion as a problem of interaction, not of transmission. Designing affective systems, then, emphasizes the support of user experience, interpretation, and understanding of emotion. Consequently, the aim of system design changes from correct recognition to providing the means for users to co-interpret emotion as it emerges in interaction.

An example of this approach to affective interaction, *Affector* is an interac-

tive installation conceptualized as a digital window connecting two adjacent offices. It is designed to help communicate moods and a sense of presence. The system collects real time video footage in each office and displays it, distorted, in the other office. It allows its users to specify rules as to how the video streams are to be distorted; the rules map office conditions such as amount of light or movement to a desired distortion, e.g., color inversions or creating ghost trails of movement. Affectors' displays may be read as emotionally informative without the system actually representing or formalizing emotion internally. Indeed, an important tenet of the interactional approach to affect is "avoid[ing] trying to formalize the unformalizable" (Boehner et al., 2007, p. 284).

Importantly, just like the interactional approach to context in ubicomp, it appears this alternative casts representationist approaches as forbidden. This tenet, then, firmly locks the interactional view of emotion on the social bank of the gap and thus further solidifies the separation between the social and the dominant technical approaches.

2.3 Gap Problems

While structuring the way we understand the research landscape, the gap also structures the kind of solutions we explore in order to address research questions. Research in HCI exhibits a division of labor along the gap's lines. The social side of the gap explores concrete opportunities to close the gap through formative studies and documents the gap through summative studies (e.g., Rode, 2011). On the technical side, computer scientists, engineers, and designers propose technical approaches, design, and build systems that implement those re-

quirements. Beyond this division of labor, there has been little engagement across the gap. For example, a citation analysis of papers in CSCW shows a schism within the field that neatly follows the social-technical demarcation (Jacovi et al., 2006). This suggests that research conversations tend to proceed in parallel on the two sides of the gap, rather than engaging the other side more deeply. Moreover, the lack of engagement between the two sides shows that efforts to narrow the gap are conducted independently from the either side of the divide, rather than collaboratively.

On the technical side, closing the gap has focused on giving technologies a better understanding of their environments so that they could sense better and respond in a fashion compatible with the social requirements. Primarily, this has involved increasingly sophisticated computational representations of human environments. As discussed earlier, sensing technologies in ubicomp coupled with powerful machine learning techniques give systems the means to infer from sensor input what the users are doing, i.e., human action. Such approaches have shown promising results in simplified environments, such as laboratory environments. Alas, in spite of significant research efforts to date there are no compelling examples of such technologies designed for realistic settings. While research on the technical side argues that more time and effort is needed for the current approaches to be fruitful, from the social bank technical efforts to narrow it are met with suspicion. This is because relying on formal representations as determinates of human attributes is at odds with the emergent nature of human activity with respect to unfolding contingencies in the environment (Suchman, 1987). More broadly, for technology to be socially sensitive it would require a presence to the unfolding of events which is not available to machines. In other words, systems would have to demonstrate human-level understand-

ing of the social setting (e.g., Leahu et al., 2008b). Needless to say, this path is problematic at best; in AI this problem is called AI-completeness.

From the social bank of the gap, alleviating the problems associated with the gap can be thought of as ‘conquering the gap by the social.’ Studies typically reveal the nuanced and idiosyncratic nature of human environments. These findings are often contrasted to the ‘one size fits most’ approach typically embedded in systems designed for such environments. As a result, social scientists highlight limitations of computational approaches to address the subtleties of social life. Consequently, it appears they put forward social fidelity — supporting existing social practices — as the main goal for system design. The effect is two-fold: first, the gap is (re)created as systems can always be shown to be rigid and insensitive to social nuances; second, expectations of social fidelity channel research efforts towards maintaining the social status quo, and thus away from potentially innovative solutions which would necessarily alter existing social practices by introducing new computational systems¹. Thus, while enormously helpful in depicting the nuanced terrain in which technology participates, efforts on the social side may appear to be conservative, as opposed to innovative, from a technological perspective.

As we can see, frustrating limitations and side-effects emerge from current efforts to bridge the gap. Technical efforts to close the gap are quickly overwhelmed by severe limitations in everyday use. In spite of advances in providing systems with more information about their users and environments, such systems seem do not understand more about what is important in those set-

¹Of course, there are notable exceptions, not least Dourish (Dourish, 2001, 2004) and Boehner et al (Boehner et al., 2007). My concern in this chapter was to present a broad and clear view of how HCI perceives and does the gap, which would allow the reader to follow the arguments in the following chapters. More nuance will emerge in those chapters.

tings. From the technical bank, the social's efforts are perceived as painting images of technology as socially insensitive and thus harmful if left to its own devices. Regardless of the perspective taken, seen through the socio-technical gap lens, current research directions appear to be stuck on either side of the gap.

Yet, both sides has something to contribute. For example, the technical side — through advances in sensing as well as statistical inference technologies — can automatically collect, organize, and analyze large data sets obtained in use environments which may contain useful information. Social and cultural insights into human environments may suggest ways in which such information may be used in humanly meaningful ways, e.g., by clarifying how people engage similar information in the broader context of their lives. As such, the two sides can be seen as opposite sides of the same coin. The challenge is to put them in productive contact, rather than seeing them as orthogonal or mutually exclusive.

A number of real world interactive technologies — such as smart phones, GPS navigation systems — show no signs of being affected by the gap. Such technologies appear to be successful both from a technical and social perspective: they manage to escape their limited technical abilities to support social practices of navigation that predate their introduction and are fluidly incorporated in people's new practices. These examples suggest that it might be possible to move beyond the limitations associated with the gap. Importantly, successful technologies at large show that artifacts — including digital — are experienced unitarily in the world. However, currently we conceptualize the gap as “real” and “inherent” (Ackerman, 2000). But if the gap is real, is it possible to do better than attempting to bridge the gap?

CHAPTER 3

THERE HAS NEVER BEEN AN INHERENT SOCIO-TECHNICAL GAP

The analysis in chapter two suggests that a divide separates sociological understandings of lived phenomena — such as human action and emotion — from the way these phenomena are approached and operationalized within computational systems. The nature and origin of this gap constitutes the focus of this chapter. HCI, I argue, has come to see the socio-technical gap as inherent in the world. This chapter opens with an overview of how this came to be and how HCI as a discipline has theorized and positioned the separation between the social and technological. Following that, I examine the gap in the context of the broader philosophical assumptions that it rests upon. This will allow us to understand the problem more deeply — why we feel stuck on either side — and its origins. I will argue that the current view of the gap stems from the underlying philosophical foundation. A different perspective into the gap emerges when one examines other philosophical foundations that do not rely on the same assumptions that make possible the gap. I overview one such alternative, Karen Barad’s agential realism. Finally, I show this alternative offers novel ways of theorizing the gap and new ways to approach it.

3.1 The Gap

In the 1980s a number of social scientists started talking about mismatches between social insights into human action and human environments and the way these are operationalized within technologies. Perhaps the most well-known example is Lucy Suchman’s *Plans and Situated Action*,¹ which locates the source

of limited (from a human perspective) interactional capabilities of technology in the fundamental asymmetries between human and machines in access to and understanding of the world (Suchman, 1987). Work in this vein had a particularly strong impact in HCI, where it led to the birth of a strand of HCI dedicated to further investigating asymmetries between people and technology and ways to overcome them. The new area, called Computer Supported Cooperative Work (CSCW), focused primarily on technologies for the workplace and work practices. In time, its interests have extended to include collaborative practices more broadly, which these days include most aspects of people's lives, e.g., social networking and online knowledge repositories, such as Wikipedia.

In this context, researchers began talking about a social-technical gap. In his call to position the gap as the overarching intellectual challenge of CSCW, Ackerman presents how researchers came to think of a gap between "what we know we must support socially and what we can support technically" (Ackerman, 2000, p. 180). His paper summarizes 10 years of work in CSCW detailing social aspects that are central to people's practices. Following Ackerman, these include: the fluid and subtle nuance of social activity; the importance of multiple, even conflicting, perspectives and goals to reaching a resolution; the central role of exceptional situations in work processes; the importance of social awareness and visibility of information exchanges to coordination, and so on.

These aspects are typically presented in relation to difficulties of computational systems to automatically support them and related practices. In contrast to the rich, nuanced, and idiosyncratic nature of social life, computational systems appear rigid, brittle, and limited in supporting social needs. As such, what we might call social fidelity' becomes the golden standard for technology: "To

summarize, there are no current HCI mechanisms to straightforwardly mechanize the naturally occurring, everyday social activity [] We must necessarily restrict the problem from what we know is appropriate to the social circumstances. This is the social-technical gap.” (Ackerman, 2000, p. 186)

An example is in order to clarify what is at stake. Ackerman examines the Platform for Privacy Preference Project (P3P) — an attempt to create a privacy standard for the web, by way of a protocol to handle automatically sharing information. The protocol is contrasted with face-to-face information sharing practices. He notes that people constantly seamlessly negotiate privacy in face to face communication. Dealing with this automatically means giving up most of that or else the problem becomes technically intractable or overloads the user. “To require users to do anything else than the apparently seamless change between ‘faces’ (Goffman, 1961) is to place users of P3P within a social-technical gap.” (Ackerman, 2000, p. 186). The example foregrounds limitations in technology to automatically replicate and support people’s existing information-sharing practices. It implicitly places the expectation of social fidelity on technology: the (ideal) goal for P3P is to fully support such practices. This sets the direction for what protocols in this area should do, even though it is acknowledged that they can never achieve this goal — i.e., the gap is forever present, even though we are forever driven by the goal of narrowing it in time.

CSCW emerged as a reaction to the general belief at the time that building technology is an inherently engineering endeavor, which need not worry about social matters. The seminal insight is that a technically working system is often not enough for it to be successful in the world. Crucially, the system also needs to be socially workable. An important mission for CSCW, then, is to reveal

aspects of social life that must necessarily be understood and addressed if technologies are to be successful in the world. The social-technical gap metaphor structures the research landscape: first, the metaphor highlights inherent limitations in the technology to automatically address social sensibilities; second, it sets the direction for making space for the social in technological agendas; finally, it implicitly advances social fidelity as a central requirement for computational systems.

3.2 Analyzing the Gap

Perhaps the greatest achievement of the kind of social science research into interactive systems described above is that it positioned the social as an important partner to the technical in the development of new technologies. In an important sense, studies revealing the socio-technical gap in practice played a crucial role in demonstrating that technical developments alone cannot handle the subtleties of social life. However, the positioning of the social on (more) equal footing with the technical unwittingly introduced specific assumptions about the technical, the social, and the relationship between them. My aim in the analysis to follow is to uncover some of these assumptions. Risking running before I can walk, the intuition here is that these assumptions are relevant to the recurrent difficulties and frustrations described by the gap.

Of particular importance for my analysis, the gap presupposes two entities — the social and the technical — given in advance, that engage in some sort of exchange in the world. Indeed, the gap understands the technical and the social as existing independent of one another. This view sees the two interacting, and

therefore influencing each other, but not in constitutive ways. Their relationship as depicted by the gap can be metaphorically seen as a dance: while the partners may influence each other's moves, they start as and remain distinct, independent individuals. In sum, the gap figures the social and the technical as **separate** and **distinct** entities.

This separation makes possible seeing one from the perspective of the other. For example, from the social perspective technology in itself appears inflexible, decontextualized, reductive; from the side of technology, human environments appear complex, uncertain, messy, while people seem 'noisy' and not always rational. Furthermore, seeing the HCI social and the technical as independent of one another makes possible a clear division of labor. Indeed, the social researcher has had two jobs: one the one hand, to describe nuances of social phenomena of relevance for interaction and, on the other, to highlight limitations of the technical to automatically handle and support human practices by studying technology in use. The technical researcher's job, then, becomes to support the requirements and to address the limitations uncovered by the social scientists. We begin to see that *understanding* the social and the technical as separate makes possible *doing* them as separate: the a priori separation of the two makes possible separate (thus different) views, which in turn motivate distinct courses of action.

The hope is that the technical will support the social better and thus the gap will narrow gradually. As we have seen in chapter 2, the main technical strategy for narrowing the gap is **representational**: representing aspects of the world internally in computational systems, so that they may understand and act in more socially sensitive ways. It is important to note that the relationship between the

social and technical that underlies the gap directly predicates this approach: as a separate entity, the only way for the technical to demonstrate any kind of social sensibility is to reason internally about the social. For this the social must be pulled into the technical, as it were. The technical strives to achieve this by simulating the social inside technology: for example, by constructing technical objects that stand in one to one correspondence with social objects such as the algorithmic representations of emotions discussed in chapter two.

Importantly, work detailing the limits of representational strategies, such as Suchman's (Suchman, 2006), generates different reactions on the two sides of the gap. On the technical side, engineers believe technical innovation will make possible better representations that will give computers more insight into their environments and thus more social sensibility. To exemplify what this might mean, in the case of the copier this would entail having more plans encoded in the system and better ways to choose between plans – like more sensors tracking users' actions. In contrast to this technical optimism, the social side problematizes any kind of representation of human practices. For example, writing about work practices, Bannon has argued that building more powerful representations will not help much as the fundamental problem is not just the need for richer notations and more resources, but rather the limitations of *any* model of work process (Bannon, 1995).

We begin to understand that efforts to narrow the gap have the opposite effect: they reinforce it. Social scientists identify limitations of and requirements for technology; engineers attempt to address and support these by adding more aspects to be formally represented in the system. This dynamic creates an endless mirroring effect: the social reflects the world including the gap back to the

technical side; the technical, in turn, attempts to better mirror the world, but the very methods for doing so — seen from a social perspective — introduce a further distance from it. Metaphorically, through the lens of the gap the social and the technical are like two magnets of the same polarity: the harder one tries to bring them together, the stronger the force keeping them apart.

3.3 Stepping Back

The picture that emerges so far is that of HCI understanding the social and the technical as preexistent categories. I have shown that the particular conceptualization of the social and the technical that informs the gap sets up a way of *seeing*, *knowing*, and *doing* that instead of gradually narrowing the gap has the effect of reinforcing it.

A different, related issue points to how we might transcend the gap. From the perspective of the gap, apart from not being able to fully support social requirements, technical innovation may bring about unanticipated (potentially undesirable) changes in social practices. This has to do with the fact that introducing novel technologies may have effects which simply cannot be anticipated. In other words, technology can be thought as having social agency beyond that which was explicitly designed into it. This observation foregrounds aspects of the interplay between the social and the technical at odds with the separate and distinct conceptualization of the two. It points to a more entangled relationship between the social and the technical. The remainder of this chapter explores how we might understand this relationality differently.

The analysis to follow has an important philosophical dimension. This is

because, as Phil Agre has argued (Agre, 1997), philosophical arguments may shed new light on technical practice. Specifically, Agre posits that every technical field is founded upon a base of commonly accepted theories or systems of thought. Such a base embodies particular philosophical assumptions which typically go undetected by the field's practitioners. For example, Agre's *"Computation and Human Experience"* takes a close look at the field of Artificial Intelligence (AI). He examines AI's technical strategies in the philosophical framework within which AI operates and traces recurrent technical difficulties back to such foundational assumptions. Moreover, his work demonstrates that uncovering and correcting problematic assumptions may open up fundamentally different directions for technical research and may thus unlock existing technical impasses. In the following, then, I discuss philosophical assumptions bound up in the gap metaphor. I am motivated to discuss them here because they help provide a clearer understanding of the dynamics around the gap and thus may inform and inspire alternatives to handle it differently.

The majority of Western science rests upon a web of foundational assumptions about what it means to know the world. This foundation provides a framework for scientific practices of knowing by answering questions such as: who can know the world? under what circumstances? what is valid knowledge? what does knowledge correspond to? Sometimes referred to as the **tripartite separation** (e.g., Barad, 2007; Verran, 2001) (or the modern constitution (Latour, 1993); Suchman calls it Euro-American imaginaries (Suchman, 2006)), this web of assumptions postulates the a priori separation of knower, knowledge, and the world. In the following, I overview this foundation by way of three interconnected assumptions that enable it:

- **observer/observation independent reality.** For science to pursue its goal of understanding and describing the world, it needs to observe reality, indeed even experiment with it. To be certain that the results and insights thus obtained actually describe the world — rather than being artifacts introduced through experimental conditions — one must make sure that attempts to understand the world do not change it in fundamental ways. While it may be necessary to intervene — for example, by attaching physiological sensors to a study participant — it is important that this does not change the phenomena observed, i.e., in our example, the person's physiological measures are not affected or if they are it becomes important that we can measure this interference and subtract it, as it were, from the observations. Thus, the notion of an observation-independent reality aims to ensure that studying or observing an event or phenomena does not interfere with it in ways that cannot be accounted for. Consequently, the non-interference notion localizes the knower outside the phenomena studied and the knowledge inside the knower's mind, in the words of Helen Verran: "Knowing is located in the minds of removed, judging observers [...], who formulate knowledge" (Verran, 2001, p. 34).
- Related to the existence of a world independent of our investigations is the belief that the world is a collection of things — e.g., objects, people, events, etc. — that exist independently of each other. This is not to mean that things do not interact and possibly interfere with each other; rather, it means that their very existence does not depend on other entities, i.e., in some sense, they are thought to be given in advance, a priori. This implies that each such thing has intrinsic boundaries and qualities: irrespective of the presence of other things, boundaries (e.g., spatial and temporal posi-

tion) and properties (e.g., color, speed) can always be observed, i.e., these attributes belong to the thing, they are intrinsically linked to it. The belief that things exist and have attributes in themselves, independent of other objects or subjects, forms the basis for a metaphysics. Barad, for example, calls it a “**metaphysics of individualism**” (Barad, 2007, p. 107), while Verran refers to it as a metaphysics of separatedness. This belief makes possible certain ways of knowing: if abstract qualities are considered to exist a priori of a human knower or other objects in the world — i.e., they are inherent in the object — when we come to know them we don’t create them, we simply describe what has been there all along.

- So far these assumptions position the knower as separate from the world and the world as a collection of things. A final step completes the circle to create a framework through which valid knowledge of the world may emerge. This step is achieved through an assumption — sometimes named **representationalism** (e.g., Barad, 2007) — pertaining to the correspondence between the world and our knowledge of it, between representation and represented. According to Barad, its origins go back to Ancient Greek philosophy, although it only gets cemented into Western ways of thinking with Descartes (according to him, representations are more readily available to our thoughts than reality). As a consequence of this assumption of correspondence, “[k]nowledge is representation of abstract or ideal categories.” (Verran, 2001, p. 34). This third assumption, coupled with the previous two, places knowledge as the correct representation of an independent reality. This is to say that knowing the world necessarily must come from detached observation of a knower; detached, in the sense that the knower does not determine, nor influence, the entity or events of

interest. Consequently, knowing the world (e.g., deriving representations of emotion) does not change the objects investigated in fundamental ways (e.g., does not modify emotion).

Having taken what may have initially seemed a detour, we may now consider the separation of the technical and the social that underlies the gap in the broader context of Western scientific practices. As we can see, separation is located at the very foundation of said practices and thus pervades our ways of knowing and being in the world. Indeed, the overwhelming majority of scientific as well as technical practices (technoscience, hereafter) rely on some form of separation, e.g., the divide-and-conquer approach: scientists delineate the phenomenon of interest from the rest of the world so that it may be understood in itself, following which, its interactions with other objects, subjects, or events may be similarly studied in isolation. Further, the tripartite separation leads to other separations, for example based on who made' the objects of study: nature vs. culture; where the objects are located: body vs. mind, matter vs. meaning; whether the object is the original or a representation: world vs. word; and so on. We begin to understand that the separation of the technical and the social that underlies the gap is but one example in a long series of separations, assumed to be inherent in the world.

An important clarification pertaining to separation is in order. This has to do with the distinction between separation as a mode of knowing — one method of investigation among others — and separation as a way of *being*, i.e., what we take the world to be like. The problem is not so much performing separation as a means to investigate the world (i.e., matters of epistemology), but forgetting that we implicitly think the world as a collection of separate entities (i.e., matters

of ontology). This forgetfulness leads to interpreting our findings as knowledge about a pre-existing world — a world in which separation is *inherent* — rather than a world in which we play an active generative role: our ways of knowing and doing perform the separation. We begin to understand that questions of ontology and epistemology *must* be considered together: knowing of the world must be understood in the context of what we assume the world to be. To this end, Barad puts forward the term onto-epistemology to indicate that the two may not be considered separately (Barad, 2007).

I return briefly to the socio-technical gap to position it in the context of the ontological and epistemological assumptions it rests upon. Approaching system design from a strict separation of social and technical and a world-view of things-in-themselves **requires** the social to mirror the world — including the technical — and the technical to mirror the world — including the social. At the same time, the assumption of an observer-independent reality guarantees that knowing the world does not change the world. For instance, constructing technical objects to mirror social objects cannot directly alter the social object because the social and the technical are a priori distinct entities; in other words, technical manipulations do not implicitly change social reality, therefore they may be **safely**¹ performed. (For example, engineers may derive technical representation of emotions as this action cannot alter the phenomena being represented, i.e., lived emotions (e.g., Picard, 1997).)

In sum, the broader (metaphysical) assumptions underlying the gap both legitimize and prescribe the course of action that led to the gap. As we have seen, the only way to close the gap is representationalist: *the social must mirror the world (including the technical) and the technical must mirror the world (includ-*

¹I.e., this is a legitimate operation.

ing the social). Yet, instead of narrowing, the gap reinforces itself. Crucially, we may now understand that **both** the feeling of safety – solidity and validity – of approaches to close the gap and the frustration of a forever present gap are fundamentally linked to the foundation on which the gap rests: the web of ontological and epistemological assumptions making up the tripartite separation.

The conclusion that emerges is that the foundation the gap rests upon may be creating the conditions for the deadlock described by the gap. To put it differently, we may be held prisoners by our own assumptions: rather than being *inherent* in the world, the gap may be an *artifact* made possible by the very ways we have come to see the world. If we are to transcend the social-technical gap, we have to examine the interplay between what we take the world to be (ontology) and our ways of knowing the world (epistemology) to figure out if and how we may understand and thus act in the world differently. This means answering the questions: how may we understand the world if we no longer assume a priori separations? Following that, how do we act in the world in ways that do not rely on inherent separations?

3.4 Understanding the World Differently

In this section, I turn to a set of theories from a diverse range of disciplines – anthropology, feminist studies, philosophy of science, science and technology studies – that have closely analyzed and problematized the triadic structure and its accompanying separations and have offered alternatives. These alternatives have sought to provide accounts of the world that do not assume such separations as inherent. For example, Bruno Latour writes about the mutual

constitution of humans and non-humans (Latour, 1993); Karen Barad examines the entanglement of meaning and matter in quantum mechanics (Barad, 2007); Helen Verran describes hybrid practices of tallying and measuring that combine two numbering systems, the English and the Yoruba, initially theorized as incompatible in any meaningful way (Verran, 2001). These readings and others provide pointers to or full-fledged theories outlining different ways to understand the world that do not rely on simple dichotomies and convenient separations. I build on these theories in the remainder of this thesis to show how we rethink the gap.

Here, then, I outline an alternative way of approaching the world, one that does not invoke *inherent* separations. Specifically, I pick Karen Barad's agential realism. Barad reworks Niels Bohr's philosophy-physics — an epistemology he derived through his work in quantum mechanics in order to settle the apparent paradox manifested in experimental observations of the dual nature (wave-particle) of light — into a full blown onto-epistemology. To develop her account, Barad masterfully interleaves insights from quantum theory, feminist theory, poststructuralist theory, and science and technology studies. Her work resolves important intellectual, scientific, and experimental deadlocks in the fields she draws upon. As such, her onto-epistemology makes for an apt tool to tackle the HCI impasse under consideration in my thesis — at once conceptual, scientific, technical, and experimental. In what follows, I sketch agential realism in contrast to the metaphysics of separatedness presented above.

An important belief supporting the tripartite separation is that our methods of investigation do not change the world in fundamental ways. Thus, the knowledge obtained is observer- and observation-independent and may be

safely thought of as describing the world pre-measurement. As I have discussed earlier, this belief rests on (1) the inherent separation between knower and the world (the Cartesian separation between subject and object) **and** (2) the possibility that we can track and account for the changes to the world introduced by our very act of investigating, and then adjust the knowledge accordingly.

Bohr shows that we may not assume an *inherent* separation between object and subject or, more generally, between object and agencies of observation. Bohr proved this point in the context of the measurement of an electron's position and momentum. To give a sense of the logic of his experiment to broader audiences, Bohr did so by way of analogy to a more mundane scenario: a person holding a stick in a room with no light. Let us consider first the two mutually exclusive ways in which a person may hold the stick: either holding one end of the stick rigidly so that s/he may sense obstacles and surfaces; or, in a loose way so that s/he may examine the stick (shape, length, etc). In the first case, the stick is part of the agencies of observation – of the experimental set up, as it were; in the second case, it is the object of investigation. Of course, the person may be holding the stick in a way that combines' these too, i.e., semi-rigidly, in which case the stick is at once object and part of subject (i.e., agencies of observation). This in-between way of holding the stick strongly disproves the *inherent* subject-object separation: the stick is at once part of both. Subject and object are inextricably entangled. Thus, in this case, the boundary between the two where the object ends and the subject begins is *indeterminate*.

Yet, when the person holds the stick in one of the two seemingly mutually exclusive² ways, the boundary *becomes* determinate. Importantly, the outcome

²This is at best an approximation: when the person holds the stick rigidly, s/he still feels the part of the stick that s/he touches; when the person examines the stick by holding it loosely, at the very least the person knows that the stick does not touch anything else or rather that it

of this determination is not fixed a priori. It depends on the material set up: in one case, the stick is the object of inquiry; in the other, it is part of the agencies of observation. It follows that the boundary between object and agencies of observation may be *indeterminate* in the absence of a *complete* specification of the experimental setup, i.e., that we know ‘a person holding a stick in a room with no light’ does not clarify the subject-object boundary. In other words, the distinction between object and agencies of observation cannot be assumed to pre-exist the experiment or the observation: **the boundary between object-subject is a priori indeterminate**. This inherent indeterminacy may only be resolved locally and temporarily — in the experiment — once all the specific conditions are in place. Material practices, then, are part of the conditions of possibility for determinate outcomes and boundaries. This example suggests a crucial aspect about the nature of separations: instead of being inherent in the world, they are *performed*. Later in the chapter, I engage this insight more fully.

So far, we have seen that separations cannot be taken for granted. Material arrangements (humans included) play an active role in the way the world is cut: what counts as object and agencies of observation cannot be settled a priori. While we may not know beforehand which side of the experiment the stick is on, the stick nevertheless exists prior to the experiment. The traditional metaphysics of individualism, or thingness, understands objects as having attributes, such as size, position, speed, and so on. In that worldview, these properties pre-exist our interactions with the object and they have determinate values that pre-exist our investigations. But if the boundary between the object of our investigations and the means to investigate it is indeterminate, the nature of measurement and measurement interactions must be clarified. In particu-

touches air.

lar, when we come to know something about an object (e.g., its length), can this knowledge be attributed to the object itself? The traditional metaphysics answers this question in the affirmative provided that we can account for measurement interaction. Specifically, if performing the measurement changes the quantity measured, we must be able to measure this change and adjust the outcome accordingly.

In an important sense, we may not be able to account for measurement interactions. Barad describes Niels Bohr's measurement interactions experiments in quantum physics. Simply put, Bohr shows that measurement interactions are indeterminate because accounting for them requires mutually exclusive experimental conditions: in order to measure the change introduced by a measuring device (or technique) we would have to be able to perform the measurement without the device, which is impossible. This is best understood by way of a concrete example: suppose we want to measure a person's heart rate. For this, we need a device that records the heart rate, for example, by way of electrodes placed on the person's body. Yet, the presence of the sensors itself may trigger changes in the heart rate. If we would like to measure this change, we would have to measure the heart rate without intervening, which is not possible³.

We begin to understand that the observer-independent knowledge presupposed by traditional metaphysics is not possible. On the one hand, the boundary between object and agencies of observation are indeterminate, rather than inherent; moreover, while it may become fixed locally and temporarily through specific material arrangements (e.g., experimental setup), different arrange-

³Alternatively, one may introduce another device to measure other changes in physiology that may help determine the measurement interaction introduced by the electrodes, say changes in skin conductance; however, an additional device may introduce other measurement interaction, so we are back where we started, namely with a measurement interaction we cannot account for.

ments may fix the boundary differently. This means that the knowledge we obtain cannot be attributed to an a priori, observer independent world, as in the case of the tripartite separation, but only to a world in which we (the observers) play an active, generative role. In other words, following Barad and Bohr: we are always part of the phenomena we are trying to understand. We simply cannot detach ourselves from the world.

By now, readers not accustomed with arguments questioning traditional metaphysics may experience a sense of losing ground⁴. They may ask: What about objectivity? Can we say anything at all with any kind of certainty? And if so, is all knowledge relative? The remainder of this chapter clarifies these questions. I want to begin by giving the reader a sense that objectivity is still possible in this alternative worldview, however the very meaning of objectivity must be understood differently⁵. The arguments above do not suggest that all knowledge is subjective. Indeed, different subjects performing the same experiment may reach the same measurement, so the knowledge is objective in the sense that it is not subject dependent. Yet, the knowledge may depend on the specifics of the situation, such as the experimental set up necessary to obtain the measurement/knowledge. This measurement/knowledge then cannot be assigned to a world that pre-exists our interventions; rather it is knowledge that necessarily includes our interactions in the world. Importantly, what Barad (following Bohr) suggests here is that the world is whole, rather than us and entities outside of us. In other words, we are of this world, rather than in a world that

⁴This is a good thing; in fact it is necessary to leave solid grounds in order to arrive at new ones. This paragraph aims to give a sense that solid ground awaits ahead, although not solid in the old way — i.e., not fixed and given in advance — but only temporarily fixed and thus always in flux.

⁵Indeed, according to Barad, the question of objectivity was an important (if not the most important) driving element behind Bohr's philosophy-physics. His efforts were motivated by his belief that it is possible to do science and maintain some sort of objectivity even when one does not assume an inherent separation between object and agencies of observation.

is located outside of us. The agential realist worldview, then, recasts **objectivity** as taking ourselves as integral to and thus inextricable from the world. This is in contrast to the traditional metaphysics which in the search for knowledge strives to remove ourselves from the world or to discount our presence in it⁶.

In the traditional worldview, the presumed inherent separation of knower, knowledge, and world makes possible representationalism as a way of knowing the world: the knower derives observer-independent knowledge which represents — stands in direct correspondence to — entities in the world. However, Bohr's work shows that local material arrangements contribute in an important sense to how and what knowledge emerges. The inherent indeterminacy between object and agencies of observation, i.e., world and knower, and the differential resolution to this indeterminacy through specific material configurations point out that knowledge of the world does not pre-exist us waiting to be discovered: it is not free floating, as it were. Rather, different doings result in different material arrangements and thus condition the knowledge produced. That is, knowing does not come from standing at a distance and representing but rather from *a direct material engagement with the world*. (Barad, 2007, p.49, original italics).

Barad pushes this insight to understand which of our assumptions need to change and how to do so. The outcome is an alternative foundation for knowing the world. First, **if our ways of knowing no longer rest on representationalism, then what does it mean to and how can we know the world?** Barad draws on the notion of performativity as an apt alternative to representationalism. De-

⁶Moreover, this profound reworking of objectivity highlights the fact that any notion of objectivity is necessarily entangled in a conceptual frame of reference; this realization deconstructs the universal aspects entailed in the traditional (or romantic (Galison, 1995)) notion of objectivity.

veloped by Judith Butler, a philosopher and feminist scholar, performativity emerged in reaction to representations of women's identity claiming to capture the essential features of such a presumed identity (Butler, 1993). In a nutshell, Butler's gender performativity questions that such features are inherent and that a universal and fixed women's identity exists. Butler shows that rather than being inherent in biological differences, gender differences are the effect of certain repetitions, in other words they are enacted or performed. Rather than pre-existing as such, then, women's identity emerges through socio-material practices.

In the broader context of practices of knowing, then, performativity tells difference from the perspective of how it emerges rather than understanding difference — and thus separation — as given. More specifically, it examines actions and doings in the world that enact differences of interest. Performativity recasts difference as achieved through material and discursive practices, rather than pre-existing those practices (i.e., being inherent in the world). Reacting to representationalism, performativity shifts the focus from issues of correspondence — representation of pre-existing attributes and entities — to issues of doings, actions, and practices that achieve attributes and entities of interest. Importantly, performativity implies that because practices continuously enact difference, changes in practices may lead to different outcomes, i.e., to reworking differences.

Second, if the human may no longer be thought of as a detached observer, then what is the role of the human and human ways of knowing? How should we think of the human with respect to the world and ways of knowing? Barad displaces the human from the center of the world and extends performativity

to non-humans. First, if we are part of the world then our actions, including practices of knowing, such as scientific practices, are examples of engagement in the world — part of the world engaging with another part of the world — rather than external interventions on the world⁷. This means that knowing, being, and doing can no longer be thought as separate, rather they entail one another. Moreover, this means that humans and human ways of knowing should not be given special treatment, they should not be considered foundational. In other words, humans should not be privileged from an ontic, nor epistemic perspective. To show how we might approach being and knowing in the world from this posthumanist⁸ perspective, Barad reworks Butler's performativity (Butler, 1993) into **posthumanist performativity**. The focus remains on actions and doings as knowing necessarily entails being in and engaging in the world, however knowing is no longer an exclusively human affair. Performativity is extended to all of nature, not just humans: in Barad's view nature continuously performs itself through actions in the physical world. Thus knowing is rethought as part of the world making itself known to another part of the world through specific material engagements.

Third, **what about practices specific to humans, like ideational or conceptual practices?** Barad's posthuman performativity reworks the nature of the relationship between matter – physical entities, such as electrons, atoms, bodies, computers — and meaning — concepts, theory, discourse, symbols, or what we typically think of ideational constructs, more broadly. Once again, drawing on Bohr's work, Barad makes the point that thinking is not separate from the material world. Bohr shows that for an ideational concept, like velocity, to be

⁷This move also serves to collapse the nature-culture dichotomy.

⁸Humanism considers the human to be exceptional; broadly speaking, posthumanism sees the human as part of nature and therefore a natural phenomenon that needs to be accounted for within nature, rather than outside of it.

well defined (to be intelligible, i.e., unambiguous) it requires a specific material arrangement that makes possible measuring it. In other words, the material and the conceptual realms are not separate, rather they entail one another. Furthermore, some of the experiments in quantum physics surveyed by Barad point to the intertwined nature of physical and conceptual practices. For example, it is now possible to perform *physical* experiments to test different *metaphysical*⁹ positions. (In an important sense, these experiments prove that no boundary can be taken for granted: e.g., quantum physics redraws the boundary between the physical and the metaphysical gets redrawn every time an experiment that was so far possible only as a thought experiment is physically realized). Therefore, Barad understands any practice to be a material-discursive practice. This understanding does justice to the mutual entailment of physical and conceptual matters: "The relationship between the material and the discursive is one of mutual entailment. Neither is articulated/articulable in the absence of the other; matter and meaning are mutually articulated. Neither discursive practices nor material phenomena are ontologically or epistemologically prior. Neither can be explained in terms of the other. Neither has privileged status in determining the other." (Barad, 2003, p. 822) The point here is that the entanglement of conceptual and material practices makes possible a common ground between human and non-human practices: the material world, i.e., physical reality. Thus, this reworking of the relationship between matter and meaning dissolves the material-conceptual dichotomy; this dichotomy, then, also turns out to be linked to our traditional pre-understanding of the world and the way we enact it through our practices of knowing and doing, rather than being inherent in the world.

⁹Literally, that which lies beyond the physical.

Pausing to recapitulate, Barad's agential realism represents an alternative worldview that is consistent with what we currently know about the subject-object dynamics. Not only does it not assume an inherent object-subject separation, it takes the inherent indeterminacy of this separation at its core, explores the logical consequences of experimental observations pertaining to the subject-object dynamics, and proposes a different worldview that is consistent with those consequences. This worldview consists of:

- Performativity: a shift from things-in-themselves to processes: examining how certain outcomes are achieved in the world, rather than what they are/represent.
- Posthumanism: removing the human from the center of the universe and repositioning her 'back' within nature. This means human practices, including scientific practices, are understood to be themselves natural processes.
- The inextricably intertwined relationship between the physical and the conceptual figures any concept or knowledge as entangled with a particular material arrangement: "Bohr's insight that concepts are not ideational but rather are actual physical arrangements is clearly an insistence on the materiality of meaning making" (Barad, 2003, p. 820). This suggests that concepts and meaning may not be considered separately from the material practices that give them meaning; and, viceversa, that material practices operate within a framework of discursive practices, which are themselves entangled with other material practice, and so on.

All of these tenets are important for the topic of my thesis because they provide a foundation for doing science that does not rely on inherent separations

and thus dissolves supposedly inherent dichotomies, such as that between the physical and the conceptual realms, or importantly between the social and the technical/scientific: “the fact is that the world isn’t naturally broken up into social and scientific realms that get made separately. There isn’t one set of material practices that makes science, and another disjunct set that makes social relations; one kind of matter on the inside, and another on the outside. The social and the scientific are co-constituted. They are made together—but neither is just made up. [...] The goal is therefore to understand which specific material practices matter and how they matter.” (Barad, 2007, p. 168) To this end, Barad puts forward a framework for how we may think about the world that does not rely on such problematic assumptions. This framework includes:

- **phenomenon.** Bohr uses this term to denote particular instances of wholeness: the inherent inseparability of subject and object. As the subject and object are indeterminate in the absence of a physical arrangement — or apparatus — the apparatus delineates object from subject and is an integral part of the phenomenon.
- **apparatus.** Physical apparatuses — i.e., particular material conditions — are needed to draw the boundary between subject and the object, and thus to achieve the separate entities. In other words, *concepts* such as ‘subject’ and ‘object’ become meaningful within a phenomenon only once the *physical* arrangement affecting the separation is in place. Apparatuses, then, are at once physical and conceptual in line with the co-constitution of the material and discursive, or physical and conceptual realms, within a phenomenon.
- **intra-action.** The subject and object can be thought to interact with each other within the phenomenon. But, the term interaction presupposes the

prior existence of two separate entities. Instead, Barad chooses the term intra-action to suggest the mutual constitution of entities within phenomena, as they do not preexist as such. Phenomena, then, are physical-conceptual intra-actions; their description must include a complete specification of the material arrangements.

- Components within phenomena, such as subjects and objects, get cut differently within different phenomena. Hence, they cannot be said to exist as such outside of the phenomena. In other words, they are ontologically indeterminate outside of phenomena and ontologically dependent within phenomena. Consequently, phenomena are the basic — i.e., the smallest — **ontological unit**: “*phenomena are the ontological inseparability of agentially intra-acting components*” (Barad, 2003, p. 815, original italics). In an agential realist view, then, the world is made of things-in-phenomena, not things-in-themselves as is the case in traditional metaphysics (metaphysics of individuality). As things get cut in certain ways only within phenomena, phenomena can be said to be “constitutive of reality.” (Barad, 2003, p. 817)
- **causality**. Causality is a relation between two (or more) entities. Since entities are indeterminate outside of phenomena, questions of causality cannot be answered outside of specific phenomena. This means a complete specification of the material arrangement that enacts the cut within the phenomena is needed. Therefore, causality can be established only within a particular phenomenon. The cut enacts a local causal structure within the phenomena: it delineates ‘effect’ entities and ‘cause’ entities; e.g., in the case of a physics measurement, the cut delineates “measuring agencies (effect) and “measured object (cause) (Barad, 2003, p. 815).

Consequently, it may not be assumed that a causal relationship between entities carries to other phenomena, as the boundary between and thus the entities themselves may be formed differently.

- In contrast to classical metaphysics, agency is not a quality that belongs to a particular object or subject. Rather, in agential realism, agency is in some sense the dynamism of the universe and manifests itself in intra-actions: “Crucially, *agency is a matter of intra-acting; it is an enactment, not something that someone or something has*. It cannot be designated as an attribute of subjects or objects (as they do not preexist as such). It is not an attribute whatsoever. *Agency is “doing” or “being” in its intra-activity*. [...] Particular possibilities for (intra-)acting exist at every moment, and these changing possibilities entail an ethical obligation to intra-act responsibly in the world’s becoming, to contest and rework what matters and what is excluded from mattering.” (Barad, 2007, p. 178)

A question that remains is that of objectivity. To provide a rigorous way of doing science, this framework must resolve the question of objectivity. In the traditional metaphysics, objectivity is made possible by the assumed separation of subject and object: as the subject is seen to be exterior to the object of study, the knowledge obtained is thought to correspond to the object. Thus, in a traditional worldview the objective referent is the object itself. As we have seen, agential realism problematizes this inherent separation. Yet objectivity is not lost; rather it gets redefined to account for the inherent entanglement between object and subject – or agencies of observation, more broadly – and the local and temporary separations between the two effected within phenomena through specific material arrangements.

Agential separability – the cut constructed between object and agencies of observation by way of specific material arrangements, i.e., apparatuses – provides the conditions of possibility for objectivity. Rather than being inherently exterior to the object of knowledge, agencies of observation (or subject) are enacted as separate and thus exterior to the object, *but only within particular phenomena*. It follows that all matters of objectivity must be examined within a phenomenon. A first question that must be answered, then, is **what is the objective referent?** Given that entities become determinate only within phenomena and are thus ontologically dependent on particular phenomena, the objective referent must be phenomena, the smallest ontological unit. Further, as boundaries between entities are drawn locally and temporarily within phenomena, agential realism understands measurements, attributes, and properties to belong to entities-in-phenomena, rather than entities-in-themselves. In this way, Barad following Bohr repositions objectivity so that it acknowledges the inherent entanglement between object and subject as well as the separation between the two effected within phenomena. In other words, this account of objectivity recognizes the subject/agencies of observation as an integral part of that which is studied (the phenomenon) rather than casting it as an entity *inherently* exterior to it, as is the case in traditional accounts of objectivity.

Since the object-subject boundary may be cut differently within different phenomena, different knowledge may emerge from different cuts. In other words, the knowledge depends on the apparatus (at once material and conceptual) performing the cut. This may raise concerns pertaining to relativism. A final question that needs to be addressed, then, is: **is all knowledge relative?** First, comparing knowledge from different cuts (and thus different phenomena), necessarily entails the entire phenomenon, including the apparatus that

produced the cut. This requires a complete description of the material set up. This ties in with agential realist objectivity's orientation to "marks on bodies," in other words to specific material arrangements, as reference must necessarily be made to bodies (physical entities) in order to give concepts unambiguous meanings: "so that there exist well-defined concepts that can be used to objectively describe the results" (Barad, 2007, p. 174). So, the apparatus effecting the cut and the material-conceptual constraints it introduces (what is included *and* what gets left out) must be taken into consideration.

Second, a comparison between outcomes from two or more phenomena necessarily entails a criterion. This is itself an apparatus and it produces a cut within the phenomenon that entails the phenomena being compared. This resonates with what Latour proposes as way of comparing or relating different entities (specifically, his work is concerned with nature-culture collectives) that appear to be incompatible. As a way of understanding possible relations between different entities (in our case, outcomes), Latour puts forth relationism: the emphasis is on mechanisms used to gauge these relations, as relations are never absolute in themselves – they depend on the criteria one uses for the comparison. The point is to measure differences and at the same time to include in the analysis and in the relation itself the 'yardstick' (i.e., criterion) used to tease out differences. (Latour, 1993). In chapter 4, I will explore what this means in technical practice.

This concludes the brief outline of agential realism, Barad's **performative** (relational) metaphysics she proposes as an alternative to the traditional metaphysics of separatedness or thingness. Agential realism dissolves the seemingly inherent divide between world, knower, and knowledge. Indeed, if the knower

is both part of the world and part of the knowledge, an a priori distinction between the three does not make sense and so these seemingly independent notions collapse into one: the world. Yet, separation and boundaries between entities are present in the world — for example, the interface can be thought of as the boundary between machine and human. In this alternative ontology, separation emerges as a result from practices (actions and doings in the world), ours as well as those of non-human actors.

The alternative onto-epistemology outlined above demonstrates that it is possible to approach the world differently: as emergent, rather than a more or less fixed collection of things-in-themselves. Importantly for this thesis, it re-frames separation as the on-going dynamism of the world – a matter of doings – and the outcomes of separations as always in flux, rather than separation as an inherent feature of the world – a matter of being (ontology) – and therefore fixed outcomes. Related and also of considerable consequence, it recasts the knower from a non-intervening discoverer, who uncovers a nature or culture that pre-exists his/her observations, to an active participant contributing to the enactment of one reality out of a field of possible realities.

3.5 There Has Never Been an Inherent Gap

In this section, I sum up the insights pertaining to the socio-technical gap that emerged from examining it against the alternative world-view presented above. Here, I focus on explain at once the illusion and the reality of the gap.

We now have the means to understand that the perceived separation of the technical and the social reveals more about our assumptions about the world,

than it does about the world itself. Concretely, when examined from the perspective of ontology (what we take the world and things in the world to be) and epistemology (how we know the world) the gap reveals itself as the combined (and entangled) effect of the metaphysical foundation it rests upon, a metaphysics of individualism, and our very methods of investigation (knowing) and building (doing) that stem from this foundation. This insight recasts the catch-22 situation described by the gap as: while we seek to bring the social and the technical closer together, our ways of knowing and doing continuously separate the two. This is to say that the gap is rooted in our worldview and it is achieved through our practices. This understanding dissolves the idea of the gap as *inherent*. The alleged fact that the socio-technical gap is a feature of the world turns out to be an artifact. As our ways of knowing and doing continuously perform the two categories as separate and distinct, it is through our practices, then, that *we* make the gap real.

While dissolving the view of the gap as inherent, this realization does not dissolve the differences between the social and the technical. Recognizing the co-articulation of the social and the technical, in other words, does not mean that there are no differences. Rather, it situates those differences as continuously enacted, rather than fixed, and therefore subject to different enactments. The ontological alternative discussed in this chapter, then, “understand[s] the nature of difference differently.” (Suchman, 2006, p. 260)

Having clarified the origin of the gap and some aspects of its nature, the question that remains is: what does the gap look like now, i.e., in Barad’s framework? The chapters to come will continue to explore this question and by grounding the explorations in technical, design, and system-in-use obser-

ventions. I return with a summative view of what the gap looks like within a relational ontology in the final chapter.

3.6 Performing HCI Differently

In the remainder of the thesis, I show what it may mean for HCI to take an alternative orientation towards matters of ontology seriously. To achieve this, I focus on one particular alternative, the one which I began to outline here; I will introduce more notions and clarifications as necessary. Hereafter, I trace how this alternative changes the very questions we ask (research priorities), the way we answer them (our process, methods), the kind of insights that emerge from our studies (knowledge), and consequently the way we build interactive systems.

It is important to note that these are first steps towards a technical practice based on a different ontological basis. At a minimum, this work shows that thinking outside of the dominant (so pervasive and entrenched that it has come to be seen as commonsense) metaphysics is difficult — partly because it has been an unquestioned background assumption of scientific and technical practices for centuries — and that getting used to seeing the world in a different way takes continuous practice and therefore time. In many ways, my work builds on my colleagues' work (in other fields and in HCI) I draw on, who have already started tackling this massive challenge. Completely removing all traces of the metaphysics of individualism, then, would be a tall order for my project. However, further addressing this ongoing challenge, in the chapters to follow:

- I show what it might mean to perform technical representations within this new metaphysical framework and contrast the process to the classical one.
- I design, build, and evaluate an interactive system from an agential realist perspective.

CHAPTER 4

REPRESENTATION WITHOUT REPRESENTATIONALISM

Representations are the bread and butter of computational technologies. The latter rely on digital representations for information about the world and in order to participate more fully in their environments. As we have seen in chapter 2, the prevailing understanding of representations in technical circles sees them as a mirror of the world: standing in direct correspondence to entities in the world. Even though they may not be perfect replicas of their real-world correspondents, technical efforts aim to produce representations that are good enough approximations. This line of thinking is supported by the belief that technical innovation will narrow the gap between representation and represented.

Others have problematized representationalism, the view of representation as a mirror of reality. Notably, Suchman's work on representations of human action has made an impact in HCI. Her work demonstrated that the prescriptive use of representations of human action by computational technologies is at odds with the way people act in the world and leads to severe breakdowns in interaction. In spite of having found a receptive audience in HCI, more than two decades after her work was originally published, very little has changed in technical practice in the way we approach representations.

The most immediate way to handle the issues pertaining to the problem of representation would be to do away with representations altogether. While it may be possible for computational technologies to interact meaningfully in their environments without using representations (Brooks, 1991) – i.e., without manipulating internal variables representing entities outside the system – not

using any kind of representations would severely limit the kind of things systems could do in the world. Rather than removing all representations, a more constructive way to deal with the problem of representation would be to rethink the relationship between representation and represented and then adjust accordingly the way computational systems engage them.

In this chapter, I begin to demonstrate a way of performing and using representations that moves away from representationalism – the assumption that representations simply reflect entities in the world. This is a tall order as technical representations are intimately associated with, if not founded on, the idea of representationalism. Yet it is important that we do so if we are serious about transcending recurrent difficulties, such as those associated with the socio-technical gap. Chapter 3 has shown that representationalism is inextricably entangled in the web of metaphysical assumptions that underlie the majority of Western techno-science. This explains in part why there have been relatively few attempts to approach technical representations differently: doing so requires not just letting go of representationalism, but also breaking out of the broader metaphysics it belongs to. In the second part of chapter 3, I have sketched an alternative onto-epistemology — agential realism — that does not rely on representationalism. The aim of this chapter is to show what it might mean to approach representations from such a worldview. I tackle this task simultaneously from a theoretical and practical perspective, i.e., grounded in concrete studies and experiments. This chapter, then, is about **understanding, performing, and using representations differently**.

4.1 Preliminaries

The main questions guiding this chapter are the following: what does it mean to represent? and, what do representations achieve? I explore these questions within a worldview of emergence which sees the world as continuously made and remade through practices of engagement. This is in contrast to the traditional metaphysics which approaches the world as a collection of things, determined a priori, and having essential attributes governed by laws external to the entities themselves. In the traditional worldview, then, representing means identifying the essential attributes of an entity, i.e., those that define it, and accurately characterizing them. But if we no longer accept that attributes are given in advance, rather they emerge and are performed in relation to other entities, then questions pertaining to practices of representing – such as what does it mean to represent? what is the objective referent of representations? how do different representations of the same object relate to one another? – need to be answered anew.

I approach these questions simultaneously from a theoretical and practical perspective, grounded in two experimental case studies. Recall from chapter 3 that agential realism sees conceptual and material matters as mutually entailing one another, rather than inhabiting separate realms. In other words, theory and practice are intertwined. In what follows, my approach will be to use the theory to understand how to (re)orient technical practice and to use practice – specific research projects – to further explicate this alternative theory in concrete terms and to work out concrete understandings and consequences ‘on the ground.’ I do so in the context of machine representations of human emotion. The core of this chapter consists of two research projects in which I perform such represen-

tations and which I use to analyze and work out what it means to represent if we take the world as emergent, rather than definable in advance.

Before describing the projects, I want to briefly revisit some aspects of theory (agential realism) that are important for rethinking representations. Recall that performativity is central to agential realism. This means the focus is on process: actions, doings, rituals whose repetition enacts certain attributes and not others. This change in focus from things in themselves to how they are achieved means representation is seen not as a static accomplishment but as the outcome of a series of doings or actions. Such doings require particular material arrangements to be in place: an apparatus is needed to cut the world in certain ways, e.g., a measuring device. The apparatus enacts a cut, that is, it generates a particular perspective by focusing on certain dimensions to the exclusion of others. What is left out, then, constitutes the perspective just as much as what is included. This is to say that representations are not innocent: they are not representations of the entities in the world; rather, they embody a particular way of looking at the world and they pertain to entities emerging through particular practices of engagement in the world.

In the analyses to follow, apparatuses play an important role. They are at once material and conceptual since concrete material arrangements are needed to give concepts unambiguous meanings and material aspects are always already entangled within a web of meanings. Throughout the description of the studies, then, I will point out how the particular conceptual orientation I take here changes the practice and does so in a consequential manner: it changes material arrangements and thus outcomes, which in turn change the theory and so on. This is to say that adopting a different worldview goes beyond simply in-

interpreting technical practice and outcomes differently, i.e., adding a different interpretation on top of a technical basis which evolves largely independent of conceptual concerns. By changing technical practice at once conceptually and materially (such as the very physical processes involved in representing emotions), the alternative worldview I adopt here deeply permeates every aspect of practice and thus has the potential to transform in important ways technical practice as well as other practices associated with the construction of interactive systems more broadly, e.g., interaction design and user studies.

As for the manner of presentation, I will go back and forth between the performative and the representationalist view of emotion, and between agential realist and the traditional worldview, more broadly. This mode of presenting will likely suit most readers as they will be familiar with the latter. Further, most, if not all, representational methods have been developed within a representationalist framework, so this way of writing allows me to focus on what needs to change in the way such representations are performed and used in order to align them with a performative view of representation.

The view of representation as always partial is not new to HCI. For example, Bannon speaks of representation as necessarily highlighting certain aspects at the expense of others in the context of representations of work processes (Bannon, 1995). Such critiques of the representationalist assumptions inherent in technical practice have been typically made from a social science perspective and to date, it is largely unclear how to engage these critiques in technical practice; that is, how should the very technical representations change to address such critiques. This is precisely the aim of this chapter. In chapter 3, I have connected critiques of representations that have emerged within HCI to broader

critiques outside of HCI, which uncovered not just representationalism as the problem, but the entire web of metaphysical assumptions entangled with representationalism. Also in chapter 3, I began to sketch Barad's alternative metaphysics that breaks away from those assumptions. Here, then, I draw on her framework for solid, coherent, and rigorous accounts of representational and techno-scientific practices, more broadly, that will allow us to move towards a rigorous technical practice constructed around a different view of representation. Having sketched a different approach to representation, at the end of the chapter, I position my contribution with respect to other alternative views of representation at work in HCI and beyond.

4.2 Emotion as Emergent and Relational

As seen in chapter 2, two conceptualizations of emotions are currently present in HCI. The dominant one treats emotion as a kind of information. It understands emotions as clearly delineable states that are internal to an individual. This perspective of emotion frames systems building as transmitting affective information between users and technologies (Boehner et al., 2007). This view of affect emerged within the traditional metaphysics, which understands the world as a collection of things that pre-exist us. Indeed, affect is taken to be a biological fact, an entity that exists in itself and can be studied by isolating it, for example through controlled laboratory studies. Furthermore, systems designed from this ontological view rely heavily on explicitly representing emotions. The representations derived substitute for user emotions in the system; in other words, they are based on a representationalist approach to representation.

A different ontological view of affect emerged in reaction to emotion-as-information (Boehner et al., 2007). Informed by historical and anthropological accounts of emotion, it posits affect as emerging in interaction. In contrast to emotion-as information, emotion-as-interaction is dynamically experienced, culturally mediated, socially constructed and experienced, and inherently ambiguous (as opposed to clearly delineable). The following points central to the interactional view of affect resonate in important ways with agential realism:

- The focus shifts from emotion as an internal state to social and cultural practices that shape and are shaped by emotion, or, to put it differently, from emotion as pre-existing interaction to how emotion emerges and is transformed through interactions. Furthermore, this view of emotion emphasizes the role emotion plays in human action and practices. This emphasis on the doings of and around emotion resonates with agential realism's emphasis that entities emerge through specific doings (i.e., performativity), rather than assuming that they are given in advance.
- In spite of the insistence to include what could be called discursive aspects of emotion – specifically, cultural and social aspects – affect-as-interaction does not see emotion simply as a socio-cultural construct. That is it does not dismiss biological aspects (bodily, physical aspects of emotion) (Boehner et al., 2007; Höök et al., 2008). Instead it sees emotion as the interplay between the two: “partially mediated through physiological signals but also substantially constructed through social interaction and cultural interpretation.” (Boehner et al., 2007, p. 278). This is consistent with agential realism's tenet that material and discursive matters are inextricably intertwined.

- Another important point that makes the interactional view of affect compatible with agential realism pertains to the subject-object dynamics. Indeed, the view of emotion-as-interaction acknowledges the role of the knower in partially constructing the object of study: “affect-as-interaction constructs its object of study as well; but it takes that construction as a necessary part of its practice” (Boehner et al., 2007, p. 288).
- Affect-as-interaction critiques the way emotion-as-information represents and uses representations of human emotions, i.e., the assumption that represented emotion stands in one-to-one correspondence to experienced emotion. Boehner et al take issue with, in my words, the representationalist stance implicitly embraced by emotion-as-information.

However, approaches associated with affect-as-interaction (e.g., Höök et al., 2008; Boehner et al., 2007) may go too far in proposing that any kind of explicit representations of emotions should be avoided: “[t]he richness of emotion in interaction mitigates against reductive representation.” (Boehner et al., 2007, p. 289). But any representation is reductive as it must necessarily highlight certain aspects at the expense of others. Furthermore, representations may still be useful even if they do demand some kind of translation and simplification of the phenomena.

As I’ve discussed earlier, the problem lies not so much in representations themselves but in the relationship between represented and representation. In other words, the problem rests in the representationalist assumptions entangled with existing methods for representations of emotions. My aim in the remainder of this chapter is to rethink representations of emotions from the perspective of agential realism, that is, to rethink representations of emotions from a perfor-

mative perspective.

4.3 Representing Emotion

From a technical and design perspective, the research presented in this thesis culminates with the design, technical construction, and user study of an interactive system for user reflection and awareness of her own emotions (see chapter 5). With the system – named Freaky – my intention was to explore what it means to make and evaluate interactive systems when we no longer rely on representationalism and a priori separations such as the one between social and technical. For this, I chose an alternative to the traditional metaphysics, an alternative that does not rest on inherent separations: agential realism.

The system uses representations which are typically associated with human emotions. To put it simply, the system gauges whether the person using it is scared based on user heart rate data. In technical parlance, the system incorporates a statistical model of fear that classifies data computed from real-time physiological sensors into two classes: fear and non-fear. Affective computing, the dominant approach to emotional interaction, understands physiology to be an information signal encoding the person's affective state (Picard, 1997; Boehner et al., 2007). Further, it constructs more complex representations of emotions – such as statistical models – to map the physiological data to the correct emotion, i.e., the emotion experienced by the user (e.g., Picard, 1997). In other words, these technical approaches see representations of emotion as characterizing the objects purportedly represented: human emotion. In the construction of my system, I want to engage some of the data sources and the tech-

nical methods for deriving representations, yet I want to do so without relying on the representationalist assumption that these representations necessarily capture the emotion experienced by the user. In the remainder of this chapter, I examine whether and how that might be possible. Specifically, I have a close look at representations of emotions from the perspective of agential realism. I do so in two related research projects: the first focuses on physiological data that might be relevant for emotion; the second examines statistical methods for classifying physiological data into emotional categories.

4.3.1 Physiology and Emotion

To date physiological data have been used in interactive systems in two different ways. First, the representationalist perspective sees physiological measures such as heart rate, skin conductance, and EEG as information channels carrying the emotional state experienced by the subject (Picard, 1997; Boehner et al., 2007). Within the classical framework, which relies on representationalism, representations based on such data aim to recover that emotional state. In chapter 2 and earlier in this chapter, I have discussed how this representationalist view of affect assumes the seamless transfer of the users' affective states from the social to the technical and back, i.e., the emotion experienced is carried by physiological signals recorded with sensors to the computational technology which processes the data and detects the emotional state experienced by the person.

The alternative computational approach around emotional experience, the one I draw on for my work because of important resonances with agential realism, embodies a different ontological view that sees emotion as emergent from

actions in the world. Systems built from this perspective have generally avoid using physiological measures to infer emotional states for fear of “formali[zing] the unformalizable” (Boehner et al., 2007). Because this view of emotion emphasizes the subjective experience of emotion and its construction in interaction, when it draws on physiological measures it does so to elicit subjective reactions and thus to aid the work of interpretation and co-construction of emotional meaning. For example, a few systems approaching emotion from an interactional perspective have used physiological measures to derive visual representations which were then presented to the user to make sense of them (e.g., Stahl et al., 2009). Such representations do not fall under the representationalist category. Instead, they see the meaning of biometric data to be open ended and thus to be made sense of by people, rather than by computers (Stahl et al., 2009).

In this section, I present the project I sketched in the introduction. It investigates further the relationship between physiology and human emotion from an emergent perspective. Specifically, the project’s goal was to understand better the process of emotional meaning-making and to explore new ways of designing affective interactive systems that do not rely on representationalism. This study provides insights into the possibilities for affective technologies to engage measurable physiological signals. It consisted of three parts: recording biometric signals using sensors, producing visual representations of the data, and co-interpreting the representations together with the participants.

Study Set Up

This study took the Bio Mapping project of the artist Christian Nold as a starting point (Nold, 2004). Nold gathers information from residents of various cities in

the form of physiological readings: residents are instructed to wander through the city, while their physiological data is recorded through the use of galvanic skin response (GSR) readers. Physiological research understands GSR to be a measure of physiological arousal, which plays a central role in many theories of affect. For example, Russell's circumplex model of affect organizes emotions in a two dimensional space, one of which is physiological arousal (Russell, 1980). Nold also continuously records volunteers' location via GPS locating equipment. After collecting the data, Nold uploads them to Google Earth, where the route taken by the participants appears mapped onto an aerial view of the location. The outcome is a map that includes each participant's route along with the GRS data plotted as a wall on the map, where at each point the height of the wall is determined by the GSR reading at that location. Connecting the individual points for a particular participant results in peaks and valleys of physiological arousal along the route.

In this project, we constructed maps similar to Nold's, using them to examine how emotional meaning arises and to probe different, i.e., non-representationalist ways, of approaching the relationship between physiology and emotion.

To record the data, we used the following equipment: GSR sensors connected to a Procomp Infinity unit from Thought Technology Ltd. and a handheld commercial off-the-shelf GPS device which recorded the location at one second intervals. We briefed the participants on the purpose and specific steps that made up the study, the data being recorded, and how the data would be used. Once the equipment was in place and turned on, each participant went on a 30-60 minute walk. There were no constraints as to where to go, however

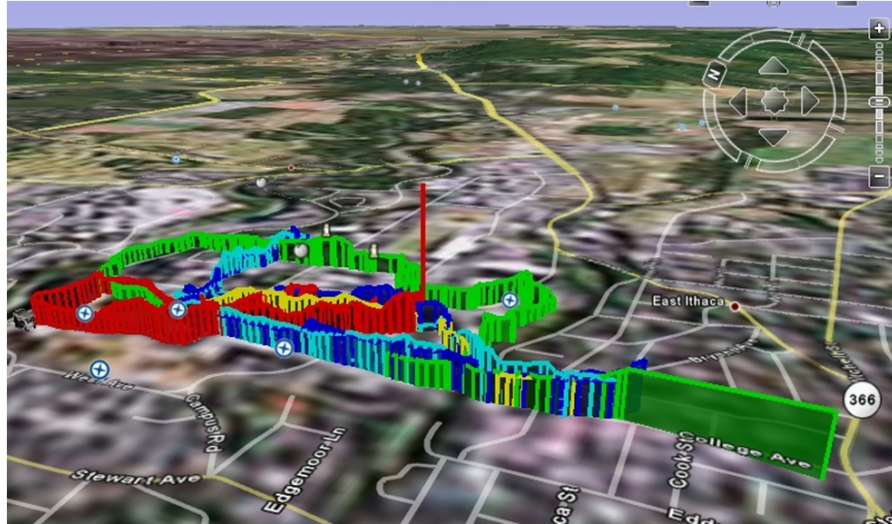


Figure 4.1: Arousal map

we suggested that they could visit places that are meaningful to them.

After the participants completed the data recording part of the study, the data were downloaded and visualized in a manner similar to Nold's. For each participant we generated a digital representation of the data using Google Maps: an aerial view of the walk with the GSR data visualized as a wall (figure 4.1). We picked Google Maps as the medium for representation.

In the third part of the study, we used the maps we constructed to explore meaning-making around emotional experiences. For this, the researchers met with each participant and gave them the map constructed with their own data. They examined the maps using Google Earth, which allows zooming in and out, changing the angle and the orientation of the visualization, adding or removing information layers such as buildings, photos linked to particular locations, etc. Following that, we discussed their experiences, the maps, and their meaning for approximately 45 minutes. The participants described where they went and

what they did, interpreted the maps together with the researchers, and imagined how other people might use them. Most of our questions were deliberately open-ended and provocative in order to invite the participants to tell stories that would provide us with a more detailed picture of the way these maps could be interpreted and the particular circumstances that make the maps meaningful.

A total of five people participated in the study: two male and three female, with ages in the 21-36 range. We took ourselves as participants (two people) and we had our friends as co-participants. This approach to recruiting participants for the study goes against established methodology of traditional HCI studies. Specifically, as discussed in chapter 3, an important condition for valid knowledge requires the knower to be outside of the object studied – i.e., a clear separation between subject and object; alternatively, if the subject-object interference is unavoidable, it should be accounted for and the knowledge adjusted accordingly, so that the knowledge pertains only to the object itself. However, the perspective I take here – agential realism – explicitly contests that the knower can ever be outside the phenomenon studied and insists that the knower must be part and parcel of the analysis and thus the knowledge produced. From the perspective we adopted, then, our decision to include ourselves as participants was meant to explicitly contest the a priori separation of subject and object.

The other participants were recruited from the researchers' social circles in order to provide a familiar context for our discussions, a setting that facilitated an uninhibited dialog around the feelings, emotions, and intimate events that were associated with the maps by our volunteers. The existing social context made it easier for us to relate to our participants by informing and pointing the conversation in a deeper way than if the relationship between researchers and

participants had been decontextualized.

In the following, I show how adopting a different ontological perspective changes how we think of representations. Concretely, I present the agential realist perspective on this study – e.g., what is represented, the ontological status of entities, etc. – in contrast to the traditional view.

Recording Data

In the first part of the study we equipped our participants with sensors and recorded continuously their location and GSR data. From a classical perspective the sensors record measurable quantities pertaining to the participants. As such, the act of measuring such data does not constitute an intervention that may alter the object of measurement: the sensors simply record something that is already there whether one measures them or not.

From the perspective of agential realism, the situation described – the participants performing a walk – is an instance of wholeness, a phenomenon. Adding the sensing equipment to the mix makes possible certain views within the phenomena that otherwise wouldn't be possible, concretely GSR and location data. In other words, each of the two apparatuses effects a cut, i.e., it produces a way of seeing the situation at hand. For example, the GSR sensor cuts – i.e., separates – GSR from the person's body. The outcome is a measurement, materialized¹ within the recording unit as a digital representation.

Measuring the GSR, then, requires a material apparatus – the sensor – to

¹In spite of the pervasive view of bits as immaterial, they are intimately tied to and cannot not exist outside of specific material arrangements, whether electric, optic or magnetic arrangements (e.g., Blanchette, 2010)

effect a cut into the phenomenon. From the agential realist perspective, in the absence of the material apparatus effecting the cut, i.e., the sensor, the GSR is indeterminate. To see this more clearly, let's examine how the GSR sensors work. The GSR is the measurement of the skin's electrical conductance, which is the inverse of electrical resistance. To measure the electrical conductance, the sensor sends a small current through the body. In the absence of the sensor there is no current to make the measurement and give local meaning to the notion of "skin's electrical conductance." To put it differently, the skin's electrical conductance does not exist a priori to the measuring apparatus being in place. In order to measure the GSR, then, the sensor *creates* the GSR: it effects a cut in the phenomenon so that GSR may *emerge* through particular intra-actions, i.e., between sensor and body.

This is in contrast to the classical perspective, which sees measuring the GSR as observing something that is already there, a property of the body that exists independent of the experiment. That is, the GSR is found by the sensor, rather than being created. The ontological status of the object of measurement (GSR), then, is an important point of difference between the two worldviews. Within the classical ontological perspective, the GSR has come to be seen as a property inherent to the body. While ontologically connected to the body, the classical perspective does not see GSR to be also ontologically dependent to a particular material-conceptual apparatus. It is an attribute pertaining to the body which exists as such, whether one measures it or not. I want to emphasize that from the agential realist perspective, the data is ontologically inseparable from the phenomenon in which a material-conceptual arrangement (which includes the sensors) enacted the separation of GSR from the body that made the recording possible. From this perspective, GSR is ontologically indeterminate outside the

phenomena – which includes the body and the material-conceptual apparatus – that is, GSR does not exist as such outside the material-conceptual arrangements that give it meaning and a way to measure it.

This ontological dependence on the phenomena must be kept in mind since the measurement makes possible manipulating the GSR data at a later time as if the data were independent of the phenomenon within which it emerged. Thus, it may appear that GSR exists independently, but this may be achieved only within the experiment, i.e., within the phenomena. Otherwise, it would be like separating dance and dancer: while it may be possible to think of them independently, in practice – i.e., when performing a dance – we cannot have one without the other. Agential realism provides a way to account for this ontological dependence by taking **the objective referent to be the phenomenon itself and the outcome of the measurement to be GSR-within-the-phenomenon**. This is in contrast to the classic ontological view, in which the GSR itself is the objective referent.

Visualizing the Data

In the second part of the study, we generated visual representations of the data. Such visual representations are a form of representation based on measurable data. From a classical perspective, they may be thought of as providing an objective viewpoint on the data. However, the agential realist view would see each representation as the outcome of a cut. Therefore, the representation is always partial as it includes certain aspects and must necessarily exclude others.

To see what is at stake, let us consider two alternatives to the digital repre-

sensation we used in the study. One such alternative is to include all the participants' data in the same representation. This first alternative representation, then, includes and extends the previous one. A second alternative departs significantly from the previous. Here, the maps would be individual and depart from the natural topology by rearranging the landmarks visited in order of increasing GSR value (an average would be computed for each landmark). This alternative representation then is complementary to the other two, i.e., the view it presents would not be accessible by looking at either of the other two maps, and vice versa. Considering which visual representation to choose, then, is not just a matter of choosing what should be represented but also of **what should be excluded from the representation**. Taking into account also what is left out from the representation problematizes viewing *any* representation as *inherently* objective.

But objectivity means different things depending on which ontological view we take. In the classical sense, objectivity means removing oneself as much as possible from the object of study. As we have seen in chapter 3, agential realism redefines objectivity as being accountable for the cuts we help enact. In other words, being accountable for the realities we help enact which necessarily entails acknowledging our (the researchers') contribution in the co-construction of the object of study. A closer look into how we may understand the visualizing process through the perspective of agential realist will make this point more clearly.

Concretely, the process starts with the raw data — bits in some form of memory or another — as the outcome of a cut within a phenomenon: the data collection, described in the previous section. This outcome becomes part of another

phenomenon within which the visualization will emerge. Recall that in agential realism, entities do not preexist as such, they exist only within phenomena – in other words, phenomena are constitutive of reality. The visualization is made possible by introducing an apparatus that effects a cut within the phenomenon. The apparatus necessarily includes the person performing the visualization as s/he picks how the data should be cut: what and how should be in and what should be out of the representation. It becomes clear that the visualization emerges from a specific material-discursive practice of engagement with the world.

The agential realist view explicitly acknowledges the constitutive role played by the person performing the representation. In contrast, the classical – i.e., representationalist – approach to representation sees the different visualizations described above as representations of the same data approached from different angles. That view may acknowledge the representer as ingenious and skilful, yet it sees the role s/he plays as a passive one: uncovering a truth that was always already there. As we have seen in chapter 3, downplaying the role of the observer (the person performing the representation, in this case) is necessary to make knowledge claims within the traditional metaphysics. In this way, the data – rather than the phenomenon which includes the researcher – ‘tells the story’.

Within the agential realist framework, exclusions are constitutive of reality. In the case of representations, this means that what is left out – or made invisible – shapes the object represented just as much as what is included. This understanding points to the fact that any representation is necessarily situated: the view it presents is taken from a particular perspective. As the feminist

philosopher of science Donna Haraway has pointed out, traditional scientific approaches promise a disembodied view of infinite vision (a 'god-trick'), instead of acknowledging that any view is always already a partial perspective (Haraway, 1988). Critically, agential realism provides a way to identify that perspective by tracing the emergence of the representation within the phenomena: the representation is the outcome of a cut effected using a material-conceptual apparatus. Examining the apparatus, then, holds the key to understanding how the representation emerged.

Agential realism sees different representations obtained from the same data – such as the ones described before – as different phenomena and thus providing different (potentially complementary or even conflicting) material and discursive outcomes, i.e., the views produced. This is not seen as a problem; rather, each such view co-performs the represented. This is in contrast to the traditional metaphysics. Because representation must be in one to one correspondence to a preexisting object, different representations of the same data must stack, i.e., they must be additive. In the classical framework, then, if different representations of the same data contradict each other, one must necessarily be false. Returning to the question of objectivity, in the agential realist framework objectivity means taking these partial perspectives into consideration including the way each was achieved and being accountable for which partial perspective(s) we make real; i.e., objectivity in this framework entails accountability for what is excluded.

Interviewing the Participants

In the third part of the study, we discussed with our participants their experiences and the meaning of the maps. A classical take on this part of the study

would have tried to identify whether any emotions were experienced by the participants and try to match those with the physiology data (e.g., Healey et al., 2010). This could have happened in a number of ways, for example by funneling user expression through a small set of pre-specified emotional categories, asking the participants to rate the felt intensity of the emotion, etc. Approaching emotion from an emergent framework, we oriented the discussion in ways that allowed us and our participants to construct narratives of what had happened and emotional meanings around that, paying attention to the work required to arrive at such meanings as well as to how articulating emotions helped our participants understand themselves and their experience. I return to explore further differences between the two ontological perspectives after I introduce some of the outcomes of the interviews.

Themes

From the discussions, we synthesized a number of themes that emerged. We presented them in a paper published in 2008 (Leahu et al., 2008a); in the following, I focus the attention to a subset which I will then discuss in the context provided by agential realism:

- **From physiology to meaning.** One of the recurrent topics was the meaning of the maps. When describing the maps themselves and their relevance, the participants described them using neutral, very literal, machine-like, non-emotionally charged descriptions, such as “just a plot.” Asked whether the mapped data could have any meaning for them, the participants struggled to take a firm position as to whether the mapped data meant anything to them or not. However, in the discussion covering what had happened on their walk or when commenting on peaks or val-

leys in the representation in the GSR data, our volunteers used the maps as mnemonic aids to recall and to explain what might have happened or to reflect why a particular point might be justified by several factors. The following participant quote expresses this transition from meaningless in themselves to meaningful when considered in a specific context:

“It maps the physical state, not emotion if it was emotion it would probably be flat. The map has no significance to me. [...] I don’t think it really tells anything about me, I mean, it does in a way, but not really. [...] Like the peak when crossing the bridge made me think whether it is because I’m afraid of heights, or maybe just because I think it’s pretty or I actually thought “I wonder if someone sees me here and think that I’m a suicide bomber and I want to jump off the bridge” and actually I chuckled a little bit.”

- **Open-endedness.** A closely related aspect to the discussion around the meaning of the maps is their open-endedness. Most, if not all, of the interpretations which surfaced during the study go beyond what is and may be represented in the map, i.e., location and physiological arousal. In other words, the volunteers engaged the flexibility of meaning afforded by the representation. Further, some saw it as an opportunity to engage differently with the environment, as explicitly articulated by one of them:

“The way we interact with the environment right now is really mediated through language and images and stuff and so maybe we should provide other ways of visualizing [the environment]; like the news are supposed to link you to the world, but “hello! they’re not!”, because they’re already interpreted for you, so I think people would be interested in [the maps] just because they are sick of being told what to think.”

- **Reflection and awareness.** The participants used the maps as a basis from

which to reflect on everyday activities and the way they may change in time or to discover things that otherwise would be overlooked. One participant wished that there would be a way to build such every-day maps over a longer period of time:

“I think it’s more interesting to do it when you’re doing more normal stuff, because then you might get to see some things that you might not realize you react to. [...]I wish we could have done a map of me over the years that I spent here. And it would be cool if it could fade in time.”

Another participant ran errands while her position and physiological signals were being recorded. In the process, she went back and forth to her office four times. The map consistently shows extremely low GSR values around the building where her office is located. She notes:

“Well, that’s unexpected! (laughs) This is pretty funny Can you blame me?!? Damn that’s pretty low I probably looked inside and I was like “Oh dear!” [...] [the arousal level] drops to the ground — I was probably thinking “Urgh papers boring (laughs) (looking at the map:) And then as soon as I get away from the building, it’s back up!!!”

When discussing allowing other people to examine their maps, or wanting to see other people’s maps, the participants mentioned significant others and close family members. Here, the maps were perceived as an opportunity to express, project, and talk about feelings:

“I would have [my boyfriend] try it out! I would just be curious to see if we would react in a similar fashion to places, although it’s so abstract, but still it’s really cool. I would like him to take a walk with it and then give it to me! I think it would be great for him, it would be such a great tool. This would be a great tool for people that have a hard time verbalizing their feelings, not that I’m saying my

boyfriend has a hard time doing that (laughs) [...] but yeah, it would be great as a reflection tool for people. I would be interested in seeing my mom's. She is not a very emotional person, but you know, she does have emotions, she just doesn't talk about them and it would be cool to see her, you know, dealing with life... you know, maybe it would give me more insight into her states."

On truth and opportunities for engagement and contesting meanings. Collecting GSR and location data makes possible aggregating large amounts of the data and visualizing the aggregates – for example, average GSR at each location – for entire cities or specific neighborhoods. One such example is Nold's "emotion mapping" project which published collective emotion maps of cities such as London and San Francisco. Through the aggregation of personal, subjective, responses to places, trends and patterns among collectives begin to emerge. To this end, he states that these maps "are packed full of personal observations which show the areas that people feel strongly about and truly visualize the social space of a community" (Nold, 2004).

For our discussion, we created we created a mock collective arousal map for the Cornell campus (figure 4.2). Shades closer to red correspond to higher than normal, blue to lower than normal arousal levels. The aggregate maps generated a wide array of usage scenarios from alleviating home-sickness, to exploring and navigating a new city, to opening up new interpretative spaces at the community level.

When asked which city's map she would like to have access to, one participant unhesitantly answered: "*[my hometown]. Man, I am so homesick, I could use anything that would connect me to that place. [...] The map would help me come up with stories, even though they weren't true!*"

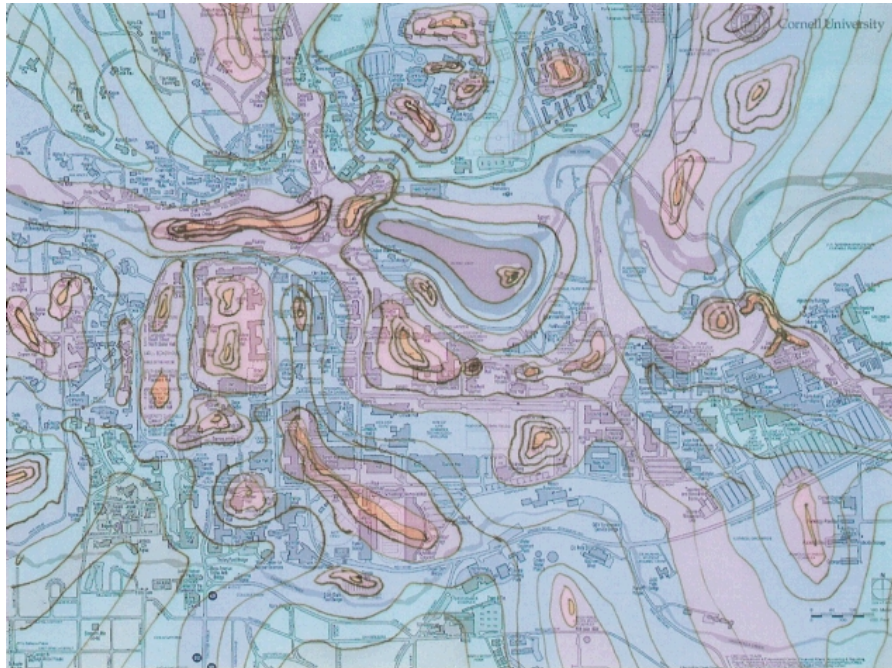


Figure 4.2: Cornell campus mock aggregate map.

An important part in the discussion around such collective maps was whether such maps capture anything meaningful at all, for example, given that minority reactions might be averaged out, i.e., disappear from view as only the average gets represented. The fact that the maps may not be ‘true’ was seen as an opportunity to provoke citizens into civic dialogue on sensitive issues (e.g., racial segregation, homophobia, high crime areas, etc.) and therefore to raise awareness and to think about what it would take to address those issues. For instance, one participant suggested using maps that apparently reinforce stereotypes: “study if some groups are overly emotional, for example if the Hispanic neighborhoods are more emotional than the Asian neighborhoods.” The emphasis here was not just on the maps, but also and more so on people’s reaction to the maps.

We drew on these and other insights into how people make sense of their feelings to explore the design space they open up. We did this by way of research through design (Zimmerman et al., 2007). We envisioned new ways of engaging people through the design of wearable affective systems – such as garments and jewelry – as well as affective displays, individual or community aggregates, to provoke rituals of reflection, awareness, and civic engagement. These are available in our published report (Leahu et al., 2008a).

Having given an overview of the insights into emotional meaning-making and how representations of physiology may contribute in this area that emerged from this study, I now turn to considering this last part of the study in the context of agential realism.

The **multiplicity of meanings** that may emerge from any of our experiences is evident in many of the participants' comments and the way they interpreted their own experiences. Viewed from the traditional perspective, different interpretations of the same event are suspicious. In that framework there may be only one correct emotional label for the experience; this has to do with the assumption of the existence of an objective truth. As such, multiple, potentially conflicting, interpretations are seen as evidence of the vagaries of people's self report (Picard 1997, cf. Boehner et al., 2007).

From the perspective of agential realism, different interpretations of the same data are outcomes of different cuts. This means a different apparatus was used to effect the cut, i.e., generate the interpretation. A closer look at the apparatus, then, will shine light on what is at stake. The apparatus includes the maps (themselves the outcome of a different phenomenon), the participant, and the researchers. Importantly, the apparatus includes the participant's material-

conceptual way of understanding the world, which includes and depends on her history, memories, beliefs, previous experiences, education, social class, etc. All of these may be relevant and may be drawn on to make sense of their own experience, as evidenced in the participants' quotes. Further, the apparatus includes the researchers' way of understanding the world; these come into play in the way s/he understands the situation and influences the specific questions asked, re-orienting the discussion, synthesizing the participant's answers, identifying common themes, etc.

Different cuts, then, may lead to different labels and narratives through which people sort themselves and their experiences out. The focus in the performative perspective is less on the label and more on sorting things and ourselves out in the understandings that emerge. In other words, the emotional category or label (e.g., happy, sad, thrilled) is secondary to the work done to arrive at it and the work the particular emotion does for achieving intelligibility in the situation at hand. So rather than being seen a problem, each cut and the corresponding label(s) or lack thereof gives a perspective. Different interpretations leading to different perspectives, then, are not competing, rather they are **complementary**. They are not the vagaries of self-report and user interpretation, but generative, creative, open-ended performances. They are not something to be done away with, rather something to be cultivated, e.g., to empower users (Höök et al., 2008).

Importantly for our new understandings of representations, the agential realist framework highlights the important, in fact, **constitutive**, role of the **context of interpretation**. As we have seen, depending of what one takes into account when interpreting the maps, the meanings at which one arrives may

be different. This is because the context of interpretation is necessarily part of the phenomenon from which particular outcomes (in this case, interpretations, meanings) emerge. In the classical framework, there is no room for including the context of interpretation: as emotions are seen to exist independent of our interaction, any interpretation necessarily must be contrasted to the 'true' emotion that was experienced at the time the data was recorded. In sum, in contrast the classical framework, agential realism allows for certain aspects of emotional meaning-making to be taken into consideration as part and parcel of the object of study: multiplicity of meaning and the active role played by the context of interpretation.

Stepping Back

Here, I review how the change in the metaphysical framework changes the way we see representation, focusing specifically on the relationship between representation and represented:

- Representing means generating a perspective into a phenomenon. This is achieved by performing a cut using a specific material arrangement (apparatus).
- Obtaining a perspective might mean *creating* the object of study – as in the case of measuring the GSR. Recall that in order to measure the GSR, the GSR must be created by sending a small current through the body. The GSR, then, is indeterminate in the absence of the material apparatus that produces a cut in the phenomenon so that the GSR may emerge as an entity within the phenomenon.

- Any cut includes certain dimensions, to the exclusion of others. Representations emerge from particular cuts, therefore, any representation is necessarily **partial**. Consequently, it implicitly **shapes** the object represented in certain ways and not others.
- Different perspectives may emerge from the same phenomenon (e.g., study) as the outcomes of different cuts. Because the cuts are the effect of introducing a particular material-conceptual apparatus, the multiple perspectives that result may be **complementary** or even **conflicting**. Each perspective redraws the object represented differently.
- The objective referent of a study is the phenomenon itself, as the basic ontological unit. The outcome of the study (e.g., measuring the GSR) is ontologically dependent on the phenomenon and is indeterminate outside of it.

Taking an emergent worldview in this study revealed some aspects around emotion that are invisible in the classical perspective:

- **Multiplicity of meaning.** Multiple emotional interpretations of the same event may be possible as each results in a different perspective into the phenomenon: the person's experience interacting with the world. Further, the context of interpretation, not just the context of the experience itself, is part of the apparatus that makes the perspective possible. In other words, *the context of interpretation matters*.
- Emotional meaning-making involves active work. On the one hand, this includes the interpretation work of the person whose experience is the focus of the work as well those s/he interacts with, including the researcher.

On the other hand, the agential realist onto-epistemology also makes possible to account in the knowledge that emerges for the people deriving representations (such as the GSR maps), in the case of our study: the researchers.

In sum, in this project I explored how we may begin to understand differently representations and emotional meaning-making by adopting a performative orientation, instead of the traditional, representationalist one. I did so by looking at relatively simple representations of biometric measures that influence and are influenced by emotional experiences. In an important sense, my study suggests that physiological data may point to potentially emotionally meaningful events; however, the emotional meaning may not be fully determined from observables (such as physiological data) as it is not static, but rather continuously in flux and contingent on particular aspects of the environment which may not be obvious *a priori*. In the remainder of this chapter, I will explore this suggestion further by investigating more complex representations of emotions: statistical models that map biometric data to emotional categories.

4.4 Cutting Emotion Using Statistical Models from Physiological Data

Representationalist approaches to emotion see physiological data as carriers of the affective state of the person whose data is being recorded. To uncover the affective information contained in such data, statistical methods are used to find emotion-specific patterns in physiological data. These patterns may then be incorporated in computational technologies in order to give the latter access to

the users' emotions. From a representationalist perspective, these patterns are objective representations of emotions. From a performative perspective, however, they help perform emotion in certain ways and not others; specifically, they focus primarily on physiological aspects of emotion: these methods foreground similarities in people's physiological data, leaving behind idiosyncratic responses to emotional stimuli, that is responses that are different from the statistical norm. In the research project I present in the following, I closely examine such methods from a performative perspective so that I may understand how we may use them in systems built from this orientation. I begin by reviewing relevant techniques for classifying physiological data into emotional categories.

4.4.1 What are Statistical Models of Emotion?

Throughout the thesis, I use 'models of emotion' or 'models' to refer to machine-inferred classifications of physiological data, i.e., unambiguous mappings from streams of physiological data points to particular emotional states. Models of emotions are generally constructed through a three part process — first, potentially relevant physiological signal streams are identified; then, physiological data are collected from participants experiencing or expressing particular emotions; finally, statistical techniques are applied to find patterns in the data that correlate to particular emotions. In the following, I give a brief overview of how the three steps are typically accomplished.

The physiological study of emotions is based on the insight that emotional experiences are inseparable from bodily manifestations of affect. The relationship between the two is complex and not yet fully understood, however it is

known to be a two-way causal relationship, e.g., a person feeling happy is likely to smile, and a person who forces a smile may start feeling happy (Laird & Strout, 2007). Researchers have uncovered a number of physiological signals that are physically measurable and can show variations with different emotions, such as the electromyogram signal, which measures the voltage from muscles such as those in the face and indicates whether and how intensely that muscle is contracted; the electrocardiogram, which measures the heart's electrical activity over time; and galvanic skin response (GSR), which measures the skin conductivity and is considered a good indicator of emotional arousal. Blood volume pressure, respiration volume, and skin temperature are also commonly used (e.g., Crosby & Auernheimer (2001), Healey & Picard (2005), Kapoor et al. (2007), Picard & Healey (1997), Picard et al. (2001)).

To obtain patterns that correspond to particular emotions, those emotions must somehow be elicited so that particular physiological data may be recorded. There is an abundance of elicitation methods used in the psychology of emotions. These methods attempt to balance experimental control with ecological validity. In general, the emphasis is on eliciting authentic as well as intense emotions in a controlled environment. Laboratory elicitation techniques include standardized emotional film clips (Gross & Levenson, 1995), collections of emotionally classified images (Coan & Allen, 2007), using emotional behaviors as stimuli (Laird & Strout, 2007), reliving experience (Coan & Allen, 2007), and masking techniques to elicit "unconscious" emotions (e.g., showing an arachnophobic person a picture which does not contain spiders but evokes them) (Wiens & Ohman, 2007). The choice of elicitation tends to be highly specific to the exact circumstances of the study and factors such as the target group and targeted emotions (Coan & Allen, 2007).

Once data corresponding to particular emotions have been collected, statistical methods are used to find regularities in the data collected which correlate with particular emotions (e.g., Healey & Picard (2005), Lisetti et al. (2003)) and which hopefully can then be used to recognize those emotions in future, unseen contexts. No one best method to identify these patterns has been found; the choice depends on the characteristics of the data and the statistical features to be used. Typically, supervised learning methods are used, meaning data points are labeled with the emotion elicited. A subset of the points are used for training the model and the rest are used for testing whether the learned model is able to correctly classify unseen data. Unsupervised methods are used as well to cluster points based on their similarity, however the link between the clusters obtained and semantic categories relevant for the classification problem — in our case, emotions — are not always clear (cf. Picard, 1997).

4.4.2 Goals of Study and Approach

In this research, my methodology was to follow the standard steps of building models and closely examine how emotion is performed in the process of deriving and testing such representations of emotion. Within the representationalist framework, the models are seen as objective ways of inferring human emotion from recordable, physiological signals. Therefore, a first research goal was to examine these representational strategies in detail and rethink them from an agential realist perspective. Specifically, the following considerations guided my work throughout the study:

- Approaching representations as performing the object represented (here,

emotion), I will keep track of which aspects of lived emotion are included and which are excluded from such representations. This will answer the question: **how does techno-science (and thus how do computational devices) cut emotion differently than humans?** Answering this question aims to provide a more detailed understanding of the ways in which human and machine perspectives of emotion are similar and different. This understanding may inform the interaction design of affective systems by allowing designers to position the models in the system in such a way that human and machine perspectives may resonate with and cross-inform one another.

- Further, I will highlight and clarify the role of the human work on which these representational strategies implicitly rely. Currently, as these techniques have been developed within a representationalist framework, the researchers' work in shaping these representations tends to disappear from view as the procedures are typically narrated in the literature in objective, decontextualized terms. Understanding the role played by human work will help us better understand the relationship between models and emotion, i.e., between representation and represented. Moreover, we may then understand how these models may be built differently, in other words, how we may generate alternative perspectives, and how these different perspectives relate to one another.

A second research goal driving this research was to gain an understanding of how to build models for Freaky, the system to be developed from a performative perspective and which I describe in detail in the following chapter. To this end, I investigated the models' performance in scenarios of interest, for example in more dynamic contexts such as the ones in which Freaky would be used,

rather than the static environments of laboratory studies. The latter are the main context in which model building is studied; the underlying assumption is that methods developed in controlled settings will work well in real environments. In the experiments performed, I have put to test this and other assumptions that underlie traditional research in this area, which are problematic from the perspective I adopt here; I will clarify why these assumptions are problematic from a performative point of view later in this chapter. Concretely, the second part of this study examines experimentally some of these assumptions, by addressing the following questions:

1. How do methods that have proven successful in laboratory settings perform in unconstrained settings?
2. Do models generalize between individuals (i.e., when tested on data from different subjects than those included in the training data)?
3. Does the context independence assumption underlying most prior work hold? I.e., how do models generalize to data from new contexts?
4. Does building models with data from more than one elicitation context improve the classification performance?

To answer these questions, I recorded physiological data in varying elicitation contexts and built models of emotion. I did so by comparing the prediction performance of models derived from relevant combinations of the data.

I begin by describing the data collection process, the data operations needed to prepare the data for learning, and then the model building procedures. I narrate these steps in classical terms, the way one would encounter them in

the literature, following which I unpack and analyze the entire process from a performative perspective.

4.4.3 The Study

To ground the study, we (myself and three students who worked on this project: Hyun Joo Noh, Steven Kim, and John Wang) took Freaky as the target application — a device that would recognize fear while the user is engaged in outdoor activities such as hiking or climbing — and therefore chose fear as the focus emotion. A primal emotion (Davidson et al., 2003), fear has a clearly experienced physiological dimension: “the experience of fear is a quick, unexpected response that can be quite intense involving increased heart rate, sweaty palms, rapid breathing, dry mouth, and so on” (Gray & Watson, 2007, p. 171). Due to these pronounced physiological changes associated with the experience of fear, the rates reported in the literature for correctly classifying fear data tend to be the highest out of all emotions. So in some sense, the choice of fear for our study is likely to maximize the chance of success for the models derived. We selected the following physiological signals for our study: heart signals (EKG), galvanic skin response (GSR), and respiration (R). The decision was based on what had been identified in the literature as most significant for fear (e.g., Sinha & Parsons, 1996; Collet et al., 1997) as well as the constraints of the elicitation settings. For example, we originally intended to record blood volume pressure (BVP) as well as GSR, EKG, and R. However, the BVP sensor must be absolutely fixed in a single place in order to pick up changes in the local blood vessel. Given that two of our data collection settings involved a fair amount of physical activity, we realized the sensor couldn’t be fixed properly and therefore ended up not

using it.

I continue by describing our data elicitation and model-building techniques, and then explain the results of the study.

Elicitation

For the study we recorded data in three different settings: controlled emotion elicitation in a laboratory setting, minimally controlled emotion elicitation on a high ropes challenge course, and neutral emotions in an indoor track field. The choice of settings reflects our commitment to explore how we may build these models for realistic environments. For example, Freaky's use context is likely to involve a moderate degree of physical activity, which does not occur in laboratory settings, but is likely to be present in outdoor settings. As such, one aspect that we had to investigate was whether physical activity would confound the models used to predict fear. In the following, I describe each of the three data collection episodes and their results.

Elicitation in the Laboratory

This first session was computer-based and took place in a laboratory. While their physiological signals were being recorded, participants were seated, watching a computer screen and operating the mouse. They played an online game in which the aim is to find one's way out of a maze; doing so requires intricate eye-hand coordination. After the person has been engrossed in the game for a while, a spine-chilling image pops up, accompanied by a hair-raising high pitched shriek. Our own experience in trying out the game suggested this is likely to trigger a fearful response from our participants.

The Ropes Course

For the outdoors data recording session, we were fortunate to have access to one of the best outdoor education facilities in the US, Cornell Outdoors Education's Hoffman Challenge Course. The centerpiece of the ropes course is a 64-foot high tower, complete with a "tree fort" platform at the top and a 400-foot double zip-line (see figure 4.3). Before the recording started, we placed the sensors and the recording unit on the volunteers. They also had to wear safety equipment (helmet and harness). The recording began at the base of the tower. The volunteers were asked to climb to the platform at the top of the tower using the ladder attached to the structure of the tower. On the platform one of the course instructors was waiting to secure the person on the zip-line. Once the safety preparations were over, the volunteer sat on the edge of the platform and launched off the platform when ready. After launch the person would travel on the zip-line roughly 250 feet and then oscillate back and forth until coming to a stop. The person would then be helped back down with a ladder. To understand when our participants were experiencing fear we asked them to talk out loud during their experience and to verbalize whatever they were feeling. Also, a member of our team tried it out herself in order to have a firsthand understanding of the participants' experience. The entire episode was videotaped to allow future examination.

Unemotional, Physical Activity Data

For training and testing purposes we wanted to have access to data points corresponding to what is often called in affective computing the "emotional neutral state": the subject is not experiencing any strong emotions. The data were recorded on an indoor track field while our subjects were engaged in a range of light physical activity (e.g., walking and jogging) alternating with periods of

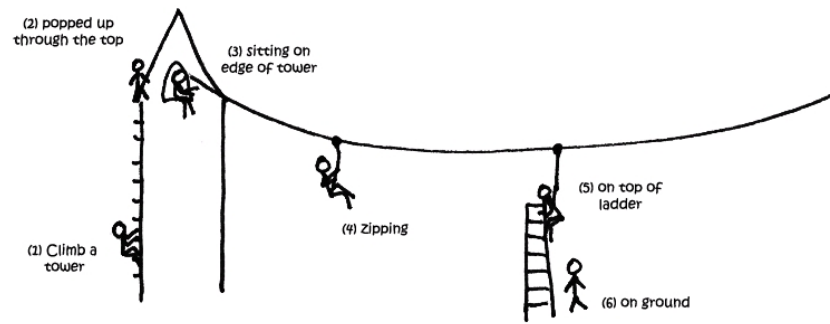


Figure 4.3: The ropes course.

rest.

Participants and Equipment

We recruited four volunteers (A, C, D, and J): two men and two women, ages 26, 31, 37 and 38. The number of participants was constrained by requiring participation in all three data collection settings and by the limited access we had access to the ropes course and its staff, as each individual recording took close to 45 minutes including set up and debriefing. We required participants' attendance to all three settings to allow us to compare within individual as well as between individual differences. The equipment used to record physiological signals was Thought Technology's Procomp Infinity unit, consisting of a decoder unit clipped to participants' clothing and plug-in sensors connected to the unit with wires. The same three sensors were used in each setting: EKG, GSR and R (sampled at 2048, 256 and 256 Hz, respectively). In the following, we discuss the results of the data elicitation sessions and then issues that arose in data collection.

Elicitation Results

In the ropes course setting each recording lasted roughly 15 minutes. The participants were asked to speak aloud about their emotional responses; the episodes were video recorded, spontaneous comments were noted with time stamps (e.g., “doing good but was really scared,” “exhilaration from climb is going down, but still nervous because of being so high up”, after coming back on the ground “still feeling a bit of after effects”), and participants were debriefed on their experience afterwards. Reviewing this data, it became clear that there were two parts where participants particularly reported fear: (1) climbing the ladder to the top of the platform, particularly the second half of the climb when the ground was getting further away and the rungs were slippery and (2) launching off the platform and down the zip-line. The recordings were affected by the temporary loss of some sensors. Particularly during the climb up the ladder the electrodes used to measure the EKG would sometimes detach or wires would be unplugged from the recording unit. This affected approximately 10% of the data points collected.

Another problem encountered was that of noise in EKG values. EKG is recorded via three electrodes placed on the subject’s shoulder and abdomen which track the electrical activity of the cardiac muscle. This sensor works fine in static settings, but once the person starts moving the electrodes pick up activity from other muscles as well. ‘Cleaning up’ the data recorded in such circumstances to remove the interference with other muscles’ activity so that the data only pertains to the electrical activity of the heart is an open research question (Kapoor et al., 2007).

As expected, elicitation in the controlled setting went much more smoothly

than on the ropes course. There were no sensor readings lost, as there was no physical activity involved: the subject was seated in a chair, watching the computer screen while the signals were recorded. Each recording lasted 2 to 3 minutes. We found that even though our subjects were anticipating a scary experience, they were clearly startled. They were quite confident in describing the experience as scary as evidenced by their reactions and commentary (e.g., C: “this scared the [expletive] out of me”). Later in the chapter, I return to contrast and analyze the controlled setting versus the ropes course elicitation.

For the third recording session — the indoor running track — each recording lasted about 10 minutes. In this setting, we also had some issues with detached sensors (particularly while jogging and running). This occurred to a lesser degree than on the ropes course (fewer than 5% of the data points were affected) as participants were able to keep an eye on the equipment and reattach sensors if needed. On the ropes course they were too involved in the experience to tend to the equipment.

Missing Data Points

As noted before, a number of sensors fell off during the outdoor data collection. In such situations, it is standard procedure to make up for the missing values. To this end, we used *k*-Nearest Neighbor, a common machine learning technique, which suggests a replacement value based on similar points in that person’s data. We experimented with different values for *k* (the number of similar points to be examined), but the results were disappointing: the suggested values did not preserve the smoothness of the signal (i.e., replaced points were highly irregular, almost random), so we concluded that these replacement values were

not realistic. Consequently, we removed the corresponding data points from our training sets. I return to the implications of this decision in the discussion following the introduction of the study set up and procedures used.

From Data to Models

Having collected the necessary data, the next step in our research was to prepare the data for the use statistical inference methods to build models of fear for the settings described above.

Features and Normalization

Machine learning (ML) techniques can more effectively find patterns in data if raw data values are normalized and transformed using statistical features which are more meaningful to the learning problem. Such transformations are essential to achieving good learning results; statistical features are generally determined through intuitions about what may be salient and through experimentation.

In our case, we first normalized the signals to account for variations between different recordings for the same person by normalizing with respect to a relaxed/normal condition, i.e., the beginning of each recording. Also, to account for variations between subjects, we normalized the signals as follows:

$$normalized_value = \frac{value - min}{max - min} \quad (4.1)$$

where *max* and *min* are the maximum and, respectively, the minimum value for that subject for that recording session (e.g., Picard, 1997; Picard et al., 2001).

The normalized signal was then compressed by way of ‘time windows,’ each

with T data points, and $T/2$ points overlapping between two consecutive time windows. Each time window was represented by one data point consisting of multiple features. Because there was little guidance for which features would be most salient for this learning problem, we cast our net widely. For each window and each signal we computed all the features found in the literature for the three physiological signals used: minimum, maximum, mean, variance, spread (maximum - minimum), sum change (the sum of individual change between consecutive points in the time window), delta between the mean of the current window and up to 10 previous windows, as well as four features based on the spectrogram (SPEC) for each signal and four features based on the power spectral density (PSD) for each signal (Picard et al., 2001). A total of 69 features were used, meaning each compressed data point had 69 values, one for each feature. For the size of the time window, we experimented with different values for T from 20 to 1000.

Labeling Points

Our initial choice for learning methods was unsupervised learning, which does not require the data points to be assigned a classification category (in our case fear or non-fear). This choice seemed a good match as it allows categories of emotions to emerge, and we hoped that fear would have extreme enough physiological characteristics that the emerging categories might correspond in some ways to experienced fear. These efforts were not successful: the clustering algorithm found the optimum number of clusters to be 5 for the outdoor data, with no meaningful correlation with participants' self-reported emotion. When forced to cluster all the points into two clusters, the resulting classification did not align with the points identified with our participants as fear. As such we decided to use supervised learning methods, which meant our points had to be

labeled either as fear or as non-fear for training and testing purposes.

All the data points obtained from the track field were labeled as non-fear. For the data recorded in the controlled setting, we knew the exact time when the person was startled, which we took to be the starting point for fear. The data before that time was labeled negative, and the data corresponding to 15 seconds after that time was labeled as positive (fear). While the beginning of the fearful event may be more or less identical to the presence of the stimulus in the video, the end of the event is not as clearly delineated. We based our decision to label 15 seconds of the data as fear on examining when the physiological data begins to return to pre fearful event values. I'd like to point out that other time frames (e.g., 14 seconds) were possible and that picking a time interval is somewhat arbitrary.

As expected, labeling the data from the ropes course proved to be quite challenging, due to ambiguity in reports from our participants for parts of the recordings. For example, after zipping and getting back on the ground the participants would talk about "feeling shaky" or experiencing "lingering effects," an experience which they related to fear but not of the same quality as, for example, the moment they launched onto the ropes. We felt that any clear classification of this felt ambiguity would be to a certain degree arbitrary, but a decision still had to be made. We decided to use this decision to orient the model's prediction bias towards characteristics that we would like the final application incorporating *this* model to have. Specifically, we would not want the application to miss an interesting event (i.e., the user potentially being scared). With this in mind and informed by the discussions that ensued between participants and researchers and reviewing the recordings, we choose to label as fear all the

points corresponding to the two parts where participants identified as generally fearful: the climbing from the base to the top of the tower as well as the entire zipping part until the person was back on the ground (about 30% of the data). The rest of the points were assigned negative labels.

It must be noted, however, that a number of other options for labeling the data were available, e.g., a conservative strategy: label as fear only the points corresponding to events when the user explicitly mentioned fear during their experience); or a more liberal approach: label as fear all the points corresponding to the climbing from the base to the top of the tower as well as the entire zipping part until the person was back on the ground — so as not to miss any points where the participants may have experienced fear; or partitioning the data into certain fear, certain non-fear and removing the ambiguous points from the training set. I return later in the chapter to a discussion of the consequences of choosing any labeling strategy on the models' performance.

SVMs

For the statistical learning part, I decided to use one of the most robust established methods: Support Vector Machines (SVMs) (Caruana & Niculescu-Mizil, 2006), using the software *SVM^{light}* (Joachims, 1998). SVMs are a supervised learning method that finds a hyperplane separating two classes of labeled data such that the 'margin' or distance to data points on each side of the plane is maximized. In evaluating the performance of generated models, three success measures are typically used in the ML literature: accuracy (acc.) — the percentage of points that are correctly labeled; precision (prec.) — the percentage of points predicted as positive that are actually positive; and recall (rec.) — the percentage of actual positive points that are predicted as positive.

By default SVMs perform a linear separation, i.e., it finds a line to partition (i.e., a cut!) the data in two; in order to achieve more complex separations data must be transformed prior to learning by a ‘kernel’ or nonlinear function which projects data points into a different dimensional space. We have trained SVMs using different kernels and decided based on the results (89.25% accuracy, 82.22% precision, and 82.81% recall) to use the radial basis kernel. A number of other SVM parameters as well as the window size were chosen based on classification results. I omit to include them here to increase the legibility of the chapter; the details can be found in appendix A.

Reflecting on Model Building Practices

From the classical perspective, model building aims to provide an objective classification of input data — physiological data — into emotional categories. This is primarily achieved by minimizing the uncertainty that might be present in the data. Examples of uncertainty in the data discussed in the previous sections include: data loss due to detached sensors (e.g., GSR), noisy data such as the EKG data affected by other muscles’ electrical activity, and perhaps most importantly, labeling the data points in situations in which there was ambiguity in the participants’ self-reports, that is uncertainty pertaining to ground truth. Traditionally, these sources of uncertainty are approached as technical or scientific problems. In the case of the first two examples, it is believed that technical and scientific innovation will overcome these problems of measurement, e.g., by inventing more robust and reliable sensing equipment or by finding better ways to replace the missing values; and, by finding ways in which the electrical signal of other muscles could be ‘subtracted’ from the EKG data, respectively.

As for uncertainty pertaining to ground truth — i.e., establishing the one, correct emotion experienced by the subject — this too is framed as a scientific problem. This framing is made possible by the ontology of separateness that underlies western science (as discussed in chapter 3). From that perspective, emotion is approached as a thing in itself (i.e., clearly delineated from other aspects of experience). Further, its scientific treatment casts it as a kind of information (Boehner et al., 2007). It follows that an unambiguous emotional state always exists — which may also be a neutral state: no strong emotions experienced. This is irrespective to whether we (as subjects or researchers) have access to it. That is, the fact that the emotional state may not always be unambiguously identified is taken as a failure on our part (either as subjects, or researchers) to access it. In other words, there always exists a clear and unique answer to the question “what is the emotional state of the user?”, but in some situations we may not have access to it. Thus, the vagaries of self-report are problematic as they seem to imply that subjects are having difficulties accessing their true emotional state. In the classical ontology issues pertaining to ground truth, then, have to do with the gap between what exists (emotional state) and what is known (our knowledge of it). The way out is to close the gap between the world and our knowledge of it by finding new approaches to elicitation and self-report that improve the accuracy of self-report or by short-circuiting self-report altogether, e.g., by replacing it with objective measures such as physiology (e.g., Mandryk et al., 2006). Put it differently, it is not that the emotion itself may be *indeterminate*; rather, our knowledge of it is *uncertain*.

The conclusion that follows is that in the classical view, the models depict an objective view into emotion. The remaining sources of uncertainty are introduced through the data and our current methods of handling emotion and the

statistical manipulation of the data. Minimizing uncertainty, then, is a technical and scientific problem for which a solution necessarily exists.

Having outlined the classical perspective on model building, I continue by analyzing the processes through which models emerge from the perspective of the relational ontology detailed in chapter 3, agential realism. Recall that, rather than seeing the world as a collection of entities that exist as such, agential realism conceptualizes the world as being continuously performed through phenomena. From phenomena, entities (objects as well as subjects) emerge: they are continuously done, re-done, and re-shaped through specific practices. The analysis to follow, then, focuses on the very practices through which computer scientists build models of emotion from physiological data. Given that agential realism includes the knower — in our case, the researchers — in the phenomena and thus sees them as active (rather than passive, in the classical view) players in the ways the object of study — emotion and models of emotion — are shaped, I begin by considering the human work that goes into modeling emotion.

The previous sections have detailed the steps necessary to arrive at computational models that classify physiological sensor data into emotional categories. Those steps highlight that building models is not simply a matter of the computer scientist pushing a button for the sensors to record the data and the algorithms to output the models. Rather, in order to work, the sensors and the algorithms require a significant amount of work from the engineers or scientists performing the study. Examples of such work include: choosing an elicitation procedure, finding an adequate setting for the elicitation, picking sensors that may be relevant and would work in the specific setting, dealing with missing values in the data, choosing how to normalize the data, preparing the data for

model learning, i.e., picking features to be computed from the data, choosing between a variety of machine learning methods, labeling the data, i.e., splitting the data into emotional categories, partitioning the data for training and testing, and deciding on one or more statistical measures by which to optimize the models (e.g., accuracy, recall).

It is entirely possible that technical and scientific innovation will remove some of these tasks that require the scientists' intervention. For example, improvements in the ways sensors attach to the body may render obsolete replacing (or removing data points pertaining to) missing values in the sensor data. However, other interventions are likely to stay. To see this, let us take a closer look at the work that must be done in order to establish what in technical circles is referred to as **ground truth**: partitioning the data points according to the emotions experienced by the user.

The classical approach for obtaining data for a particular emotion requires creating the conditions for elicitation. In this case, the participants are subjected to an emotional stimulus followed by their emotion reaction. Hence, emotions appear to be **found**. The data recording sessions typically take place in simplified settings such as the laboratory studies. Indeed, our experience eliciting data in the laboratory shows that once the stimulus is presented an emotional response from the subjects readily follows. Thus, the emotional event may be clearly identified as its start coincides with the presence of the stimulus.

However, when we examine elicitation in the wild a different story emerges. A closer look at what it takes to obtain physiological data for fear on the ropes course suggests that it would be a distortion of facts to liken this process to finding clearly defined events in the world. For example, it is not at all clear

where the emotional response starts and where it ends. Further, unconstrained experiences rarely fall within clearly delineable emotional categories. This is evidenced by our volunteers comments while on the high ropes course: while most comments were related to the anticipation of and experiences around fear (e.g., “exhilaration from climb is going down, but still nervous because of being so high up”), none of them revealed an experience that would fall easily into a clear emotional category (e.g., fear) such as the ones observed in the laboratory setting. I’d like to point out that the ambiguity that is typically attributed to self-report and to emotional experiences has less to do with participants’ experiences – indeed, their comments give us a good sense of their experiences. Rather, their experiences appear to be ambiguous from the perspective of a clear emotional category. In other words, *it is ambiguous how to categorize their experience into the standard emotional identifiers.*

And yet this ambiguity must be resolved in order to arrive at the labeled data points model training requires. As we have seen, this requires active work from the researchers as well as participants: e.g., sorting through these events in order to relate them to fear, deciding where a particular emotional episode starts and where it ends, etc.

Importantly, there are many different ways to do this — i.e., split the data into emotional categories — as highlighted in the earlier discussion of our elicitation results. Choices included eliminating some of the ambiguous data, labeling as fear only the events one is certain about (although how that is to be established is itself open ended), etc. None of the options could be considered a priori the obviously correct choice. That is, there are no external, objective criteria to decide how to resolve the labeling ambiguity.

Nor was deciding between the alternatives random. Our decision was primarily driven by the kind of events — related to fear, with a broad understanding of fear in mind — we wanted to orient the model towards. Specifically, we decided to use this decision to orient the models towards providing the information that we would like the final application incorporating *this* model to have access to. We would not want the application to miss an interesting event: a physiological response that may be interpreted as fear. The next step was to understand what this may mean on the ground, i.e., for our participants, in the physical space in which the elicitation occurred, etc. To identify episodes to be labeled as fear we used a number of sources of information from the elicitation episode: the participants comments as they were going through the ropes course, we examined the physiological data for spikes or changes in the data that and their reflections after the event, we considered the physical aspects of the ropes course that could factor into fearful experience (e.g., climbing up the lather, being high up on the platform). These situated resources allowed us to more concretely — in terms of specific physical landmarks and time frames — narrow down and articulate the kind of events we wanted the models to capture.

In sum, the process of labeling data points was driven by the kind of fearful and physiological events we wanted the models to recognize together with resources situated within and those that emerged from the context of elicitation. To put it differently, the researchers' decisions on how to resolve the labeling ambiguity were situated and contingent.

Stepping back and reflecting on the process of dividing the data points obtained from the unconstrained elicitation context, a number of inter-connected

observations emerge:

1. **Resolving category ambiguity and bias.** When considering data recorded in the wild, as I have described, there are no definitive, a priori criteria to split the data. This is because there is no unique way to answer the question that guides data labeling in the traditional perspective: “which of these events are fear?” And yet, in order to use the model building and testing algorithms and procedures the ambiguity must be resolved. Approaching model building from a performative perspective, we have used the question “what do we want the models to do?” (i.e., what kind of events do we want the models to recognize?) to guide the labeling decisions. Answering these questions meant carefully examining specific details present and emerging in the environment. In other words, answering those questions is a necessarily local and contingent process.

In classical terms, then, the models are biased. Typically, from that perspective, bias is undesirable and should be removed so that the (one) correct, objective model may emerge. My experience on the ground shows that not only is such bias unavoidable in the wild, but it is actually needed to resolve the ambiguity in a way that makes sense locally. Rather than doing away with it, we need to orient ourselves towards it so that we may actively choose which partial perspective we would like the models to capture.

2. **Curating rather than finding data points.** The picture that emerges from the process of categorizing data recorded in unconstrained settings is that of carefully selecting the events to be categorized as fear. Indeed, the process involves closely examining, delineating, and relating these events to each other and to the idea of fear that we want the models to capture.

This is not unlike the process through which a set of artifacts are selected so that, framed and positioned in a certain way, the collection may tell a story and give a point of view. Indeed, describing and explaining the criteria and the steps through which labeling was achieved contributes to constructing — i.e., shaping and molding — a point of view from which to ground the models. In sum, data points are actively curated so that a point of view may emerge, rather than being found, i.e., emerging by themselves.

3. **Fear as an ideal category.** That curating events is needed so that a particular perspective may emerge points out that many perspectives to choose from may be available. Therefore, a decision must be made as to which perspective we want to present. This is in stark contrast to the classical perspective in which elicitation aims to find instances of fear, that is, splitting the participants' experiences into fear and no-fear is simply a matter of deciding which of the experienced episodes best *represent* fear — an ideal category, that is given in advance. Yet, as we have seen, in the wild *the very meaning of fear becomes ambiguous and must be renegotiated on the ground*. Therefore, every time we decide whether a particular experience should be in or out, we implicitly make a statement about what counts as fear. In this way, we continuously redraw the boundary between fear and everything else, and between emotion and other aspects of experience. Labeling data points recorded in unconstrained settings highlights that finding *instances* of fear and defining the *category* fear go hand in hand: the two are inextricably entangled. Further, labeling in the ropes course foregrounds emotional categories as simple abstractions of complex human experiences. Indeed, as we have seen, to arrive at fear first we must form

an event — i.e., delineate it from the rest of the embodied experience — and detach it, as it were — i.e., by setting it against the rest of the recorded experience. Thus, the emotional category is revealed as an abstraction, a shorthand to refer to a complex phenomenon.

4. **On elicitation in the lab.** The elicitation procedure performed in the lab appears to tell a very different story than the one in the wild. Importantly, there is hardly any ambiguity: once the emotional stimulus is present the subject either experiences the target emotion or not, which may be verified relatively straight forwardly via open-ended self report or requiring the subject to pick the emotional category from a list and pick a number corresponding to the felt intensity of the emotion picked from the list. It may appear then that fear is simply found (or not) by the researcher.

From a performative perspective, however, we must also consider the work that goes on in creating the conditions for the subjects' experiences to emerge: the study design. When we include aspects such as the selection of a stimulus and of an elicitation context a different picture emerges. Through these decisions the range of experiences the subject may have is narrowed so as to focus on the emotion that is the object of the study — fear, in our case. In other words, such work answers the question “what stimulus is likely to evoke fear?” and “what would fear look like in this context and in the presence of the chosen stimulus?” In other words, most of the ambiguity that is present around emotion in real — i.e., unconstrained environments — is resolved *before* the experiment takes place. As such, the experiment itself may run smoothly and verifying whether the sought emotion was present is more or less a yes/no question. In a simplified setting, then, it may seem that a priori criteria to resolving any

residual ambiguity are possible, because what fear may look like in that context has already been decided before the experiment takes place.

Thus, an implicit construction of fear is achieved through the experiment design, i.e., choosing the specifics of the elicitation process. This happens to a degree in the wild, as well: clearly, we chose the ropes course because most people find the experience frightful, among other things. But out there it becomes very clear that fear is not one, clearly defined thing, but results from funneling complex experiences into an abstraction. That is, the emotional category is an umbrella term to refer to a broad type of experiences. In simplified environments, focusing only on the stimulus and participants' responses creates the impression that the experience of fear is a unitary, homogenous, and clearly delineated event.

In sum, the classic perspective on approaching elicitation in the lab invisibly constructs the experience as an instantiation of the ideal emotional category that is the object of the study. Thus, it may appear that (1) the category was there as such all along, i.e., not reshaped in the very process of designing the experimental procedure and (2) the instances were found.

Helen Verran, the anthropologist whose work inspired this analysis, writes that through the invisible construction of instances of presumably ideal categories that pre-exist us, "difference is remade as sameness" (Verran, 2001, p. 87). To see this in our context consider the fear instances obtained in the lab and those in the wild. Because the process of obtaining instances is typically narrated in the classical view in decontextualized terms, the different, specific contexts and doings — sequences of actions — through which these instances emerged disappear from view and thus what we are left with are the objects obtained instances of fear — as exam-

ples of the same ideal type. This move, Verran contends, makes possible in the classical framework comparisons between different contexts. For my project this means that studies in which very different instances — i.e., obtained in very different ways, e.g., in the lab vs. in the wild — may be treated in the same way as they have been constructed as instances of the same class

This is exactly what happens in studies of emotion classification. Indeed, statements such as “state of the art emotion classifiers identify fear with 80% accuracy; with our new technique we have improved classification accuracy to 85%” are common in technical papers. Seeing instances of fear across studies as the same allows for classification performance comparisons between different models obtained with data from different contexts. Comparisons are not a problem in themselves; rather, the way they are done is problematic, i.e., only the very tip of the iceberg is compared, in the form of classification performance. To my best knowledge no studies have been conducted to thoroughly examine to what degree these instances are similar and how similarity might be established in this case.

From a performative perspective, such assumptions are problematic, not only because the subjects’ experiences across these instances and across different contexts may be very different but also because the way the instances were cut — i.e., separated from the rest — differs between contexts. Yet, it may still be possible that the instances are similar from the perspective of numerical and statistical features of the data. However, this must be investigated rather than assumed. To this end, the data I collected allowed me to begin to ask and answer such questions in the experimental results which make up the next part of the study.

From this analysis, a couple of meta implications emerge. The classic way of narrating emotion categorization in objective, decontextualized terms makes invisible the fact that finding emotion is enacted. This has to do with the fact that to date most such studies have been performed with data elicited in simplified settings such as the laboratory. Indeed, as I have shown, in simplified settings it may appear that there is one correct, i.e., objective way. Including the process of crafting the elicitation experience in the analysis, however, reveals that the ambiguity pertaining to clearly categorizing experience is removed before the experiment takes place. Thus, emotion may be done differently in constrained settings as well; to achieve that one needs to change the experimental procedure. In contrast to simplified settings, unconstrained environments such as the ropes course bring to the fore that multiple ways of doing emotion — categorizing complex experiences into emotional categories — exist. The differential enactment of emotion is not particular to formal (i.e., scientific) practices around emotion. Indeed, the project described earlier in this chapter on representations of physiological measures highlighted that emotion does not exist as such a priori, but emerges and is continuously reshaped through our practices. The picture that emerges is that of multiple ways of doing emotion, whether approaching emotion from a scientific or lay perspective. These observations point out the fact that **all** practices around human emotion are **creative** practices.

Thus far in this project on models of human emotion, I have focused simultaneously on the process through which these representations emerge and on the object of study: human emotion. This has allowed me to notice that the process shapes the object of study and that this is not just inevitable, but also needed: work is necessarily involved when emotions are done, whether the context is everyday practice or more formal, techno-scientific practices. I now turn my

focus more exclusively to the process itself.

The outcome of studies formalizing emotion are not just formalizations of emotions, but also methods for doing so. These methods are themselves representations (i.e., simplifications, abstractions) of the situated processes through which emotion and representations of emotions emerge. From the classical perspective — which includes the majority of technical practice in HCI — these methods, i.e., the representations of the embodied practices, are seen to capture it all, i.e., all that is needed when applying the method on the ground. As I mentioned before, the vast majority of work from which these methods were distilled have been performed in simplified settings. My experience shows that in such settings it is possible to narrate the process of labeling data and training models, more broadly, in decontextualized terms. This creates the expectation, then, that the methods will work in everyday human environments simply by following *ad literam* the steps prescribed by the method. Yet, the situation on the ground in unconstrained settings is quite different: as we have seen, applying the steps prescribed by the method depends on local contingencies in the environment which give local meaning to the decontextualized steps prescribed by the method. By leaving them out from our descriptions, e.g., of the data labeling procedure, we are not just setting aside a secondary variable, but rather the very factors that drive the process in practice, i.e., those that make the abstract instructions meaningful in practice.

This insight reiterates Lucy Suchman's argument on the relationship between efficient descriptions of human actions (plans) and situated action (Suchman, 1987). What I want to draw attention to here is that representations of technical and scientific practices — e.g., methods — are not exceptions to Such-

man’s argument. Such practices are situated, just like any other human practices. Methods, i.e., formalisms based on such practices, then, serve to orient us — those who carry them out — towards certain contingencies and away from others rather than to fully specify the actions we must follow on the ground. As we have seen, the exact way in which we engage contingencies is consequential for the outcomes of our doings and may not be specified in advance. Therefore, **we need to extend our understanding of what counts as a technical practice — and the ways we describe it, e.g., in textbooks or publications — to include the ways in which we engage such efficient descriptions — the methods — to arrive at the concrete sequence of actions that enacts the method.** I return to the insights presented here in order to discuss how we may do technical practices differently, after presenting the model testing experiments and results.

4.4.4 Classification Experiments

One of the implications of the previous analysis was that in the classical framework the work done to draw the boundaries of the ideal category — fear — disappears from view. As such it may appear that instances of fear are found. Further, having established them as instances of the same category allows for comparisons among them or among representations — e.g., models — based on them. In other words, seeing the models as representations derived from instances of the same, presumably given, abstract category allows for comparisons between different studies — i.e., models obtained with data from different contexts — to be reduced to simply comparing performance metrics such as classification accuracy. In this part of the study, I aimed to examine experimentally a number of assumptions that underlie research in this area. I do so by

training and testing models using different subsets of the data set. Specifically, here are the questions that I set to examine:

1. How do methods that have proven successful in laboratory settings perform in unconstrained settings? The assumption is that once we have good methods for constrained settings, they may be applied more or less in the same way in the real world.
2. Do models generalize between individuals (i.e., when tested on data from different subjects than those included in the training data)?
3. Does the context independence assumption underlying most prior work hold? I.e., how do models generalize to data from new contexts?
4. Does building models with data from more than one elicitation context improve the models' classification performance?

To examine these questions in the context of fear classification, we trained and tested a large number of models using various mixes of the data collected. Of the 69 features computed from the signals, we generated SVM models using subsets of the top features, ranked based on each feature's gain ratio (a measure of its relevance for classification). The best results in terms of accuracy were obtained for the top 15 attributes. We used only those attributes to generate the models for which results are presented here. Also, to estimate better how these models would perform in practice, 10-fold cross-validation was performed and the averaged results are reported. I present here the subset of the results that allow me to address the questions listed above.

Recognition in Unconstrained Settings

The first research question examines the assumption that methods proven successful in laboratory settings will work well in unconstrained settings. To answer this question, we build models using the data collected in the outdoors setting. We refer to this first group of models as aggregate models: we trained on data from three people, but tested on data from the fourth. The performance metrics for the aggregate models are presented in the right side of table 4.1.

Subject	Personalized			Aggregate		
	Acc.	Prec.	Rec.	Acc.	Prec.	Rec.
A	92.3	89.2	88.4	84.9	73.6	88.2
C	88.6	81.1	66.2	75.9	43.1	17.6
D	88.4	72.9	99.6	84.6	72.9	81.3
J	91.1	85.6	90.0	69.6	56.3	51.8

Table 4.1: Comparison between personalized models and aggregate models for the ropes course data.

Previous work on models of fear had been performed only in laboratory settings. Good fear detection rates — 83% to 87% depending on the classifier used — have been reported for such settings (Lisetti et al., 2003). Our models built with outdoors data show an average accuracy of 78.7%, however, when it comes to detection of fear in the test sets (recall) the models only detect on average 59.7% of such points and exhibit high variance between subjects. Consequently, for some of our subjects the actual recognition of fear points seems to be very brittle. First, these results suggest that approaches that were proven effective in constrained settings may be significantly less effective when used for real-world data. Second, the data points to large inter-personal variations. These are not typically observed in constrained settings. The assumption underlying most research on emotion recognition is that because good normalization techniques

have been developed, models will generalize to other subjects, i.e., beyond those whose data was included in training sets. The large between-subjects variance in the results suggests that this may not be the case for unconstrained settings.

In the following, we examine whether personalized (subject-specific) models provide significant recognition improvements.

Do Models Generalize Between Individuals?

Due to the idiosyncrasies that may be present in less constrained environments, but not in a laboratory setting, we anticipated that it might be necessary to build personalized models for individuals in order to correctly classify emotions. Table 4.1 contrasts the performance metrics for individually trained models (where a subset of a person's points were used for training, the rest for testing) with those for aggregate models.

The results confirmed the expectation that personalized models offer better predictions. Indeed, average accuracy improves by 11.35% over aggregate models. Also, and perhaps most importantly from a fear detection' perspective, the performance of classifying actual fear points (i.e., those points chosen by us as corresponding to fear) increases significantly: average recall 86.2%, compared to 59.7% for aggregate models. For subjects A and D the differences were quite subtle, while for C and J the increase in performance when using personalized models was more dramatic. Some of the performance improvement of personalized over aggregate models may be illusory, as personalized models are more likely to overfit. Nevertheless, the results suggest that *in unconstrained settings it may be fruitful to consider creating individually personalized models of emotion*, rather

than aiming for models that generalize across persons.

Do Models Generalize Between Collection Contexts?

With the third research question, we investigated the models ability to generalize to new contexts. In other words, we were testing existing approaches' potential to extract patterns that are useful across collection and, ultimately, deployment contexts beyond the one that was used for training. Also, as one of the elicitations occurred in a controlled environment, we wanted to tease out further differences between techniques validated for constrained environments and more complex, everyday-like settings.

We started by using models trained with in-lab data to classify ropes course data. The average performance metrics for the four participants paint a picture of extreme fragility for the models built using data from controlled environments: more than half of the points were misclassified (i.e., accuracy 45.6%, considerably below the baseline accuracy of 70%); only 3.1% of the points predicted as fear match the fear label and only 1.9% of fear points in the test set were detected. The opposite scenario was tried as well, testing the outdoor models on in-lab data. The results were 77.7% accuracy (83% baseline), 66.6% precision and 2.4% recall, meaning very few points were classified as fear, of those about two thirds were correct. These results expose the models' lack of robustness to different contexts.

Gain Ratio

Although both the ropes course and the in-lab data sets contain emotionally tagged data points derived from the same physiological sensor measures, the classification results discussed above suggest that the structure of these data sets — from the perspective of the computational features used — is quite different. This is directly related to the assumption discussed earlier in this chapter that instances of the same abstract category, fear, may be treated as the same, i.e., from a machine learning perspective as having the same underlying statistical structure.

In order to test whether the problem was in our selection of machine learning techniques or whether data from different contexts simply do not have the same structure, we examined the gain ratio (i.e., a statistical measure for the relevance of a particular feature) for each of the 69 features used for the two data sets. The top four attributes are the same in both cases: the features computed from the EKG spectrogram. However, for the majority of the rest of the features, the gain ratio rank varies significantly between the two data sets. This means that the salient features needed to classify the points differ between collection settings, suggesting that the patterns found in data are also different. The consequence for systems using such models is that *it may not be possible to generate models that work well in a setting for which they have not been trained*. In other words, models may not generalize well to other contexts of use.

Confounding Factors

A significant difference between simplified contexts and everyday ones is that confounding factors that typically abound in the latter type of context are simplified away in constrained environments, especially in laboratory studies. For the domain of affect recognition, physical activity may be a confounding factor, as it may lead to changes in sensor readings that are not due to experienced emotion. Here, then, we examined the ability of the models to discriminate between emotion and physical activity: we used the unemotional data from the indoor track setting to test the models built.

First, we used the personalized models obtained from ropes course data to classify unemotional data. The models classified a large number of points as fear, which translated into poor average accuracy: 72% (A: 81%, C: 60%, D: 52% and J: 95%). The majority of the misclassified points correspond to physical activity (walking and jogging), while non-physical activity (resting) was correctly classified as non-fear. On average 28% of the unemotional data was labeled as fear suggests that physiological signals during physical activity look similar to those from fear elicited outdoors and the models are not able to discriminate between the two. We repeated the experiment with models trained only on in-lab data. These models had even more difficulties classifying non-emotional, physical activity data: average accuracy 53.2% (A: 60%, C: 49%, D: 48% and J: 56%). Examining the predictions uncovered that almost all of the points corresponding to physical activity were classified as fear. This is in fact somewhat intuitive: in static, controlled elicitation scenarios (e.g., in a laboratory, with the subject seated) changes in physiological measures can be attributed to affect. However, in everyday situations there may be many reasons for such changes.

Hybrid Models

As physical activity appears to be a confounding factor for both controlled and uncontrolled models, we considered using the unemotional data to ‘guide’ the models in discriminating fear from physical activity. We tested this hypothesis by deriving hybrid models, where training sets were modified to include that person’s indoor track setting data. Interestingly, the hybrid models thus obtained for the ropes course data detect on average 8% less fear points in the ropes course set. This may mean that from the models’ perspective some of the fear points are more similar to non-fear physical activity than to other fear points. When tested on the in-lab data the hybrid models detected no fear, i.e., the fear instances in that data set are more similar to non-fear physical activity than to fear points in the ropes course data. Similarly, the performance of hybrid models for in-lab data degraded even more — on average, only 1% of the fear points were recognized. These results problematize the conjecture common in technical circles that larger, user specific data sets would improve classification in situations where difficulties arise due to confounding factors (e.g., Intille et al., 2004).

Data from other sources such as accelerometers could be used to detect physical activity and thus help clarify some of the models’ confusion. Such a move would indeed be helpful for applications in which the user is unlikely to experience fear while engaged in physical activity. However, for less constrained settings such as the ropes course, knowing whether the subject is involved in physical activity does not help as fear occurs *during* and *because* of physical activity.

Summary of Classification Experiments

The aim for this part of the research was to test whether pervasive assumptions structuring research in this area hold. To this end, we tested the status quo of statistical classification of physiological data into emotional categories in a less constrained setting — i.e., similar to a real context of use — and contrasted their behavior between simple settings (similar to the ones typically used in such studies) and more complex ones (similar to realistic deployment scenarios). These results may be specific to our domain of study: fear. In brief, here are the answers to the research questions that guided our work:

1. Existing methods and approaches validated in constrained settings may be seriously challenged by everyday settings.
2. Subject-specific models may be needed in order to detect the phenomena of interest with high accuracy.
3. Models show difficulties in generalizing to new contexts.
4. Confounding factors may overwhelm models in everyday settings. Larger, more heterogeneous data sets may not improve recognition.

Fundamentally and irrespective of the ontological perspective taken, this study suggests that resources (e.g., sensors), techniques (e.g., normalization, feature extraction) and approaches (subject-independence, context-independence) that have been demonstrated in controlled environments may not be able to scale to uncontrolled settings. This insight raises questions about the a priori generalizability of classification methods validated only in laboratory settings to the environments of everyday life.

Moreover, this research suggests a tradeoff faced by system designers using statistical classification for their systems between system autonomy and broad context of use. At one end of the spectrum, if the technology is to be highly autonomous (i.e., minimal human supervision) then the application environment must be very narrow (e.g., riding a rollercoaster) so that regularities in the data streams correlate to events of interest. At the other end, if the system is to be used across a variety of contexts (e.g., smart phones) my experiments show that the models' predictions may not be meaningfully correlated with events of interest.

Finally, while these results cannot be taken as categorical answers to the specific research questions that guided the training and testing experiments, they do problematize the unexamined assumptions that informed these questions. First, they draw attention that such assumptions may not be taken for granted, but should be thoroughly investigated. While it may be entirely possible that models may be *made* to work in the ways suggested by the assumptions, this may not happen by itself — i.e., warranted by the presumed structure of the object of study, emotion, or more broadly of the world itself. Rather it may require particular ways of performing the models and emotion itself. These ways, the work required, the inclusions and exclusions that they entail, as well as their consequences must be carefully studied.

A potential limitation of our study pertains to the number of participants, $N=4$. For the domain of affect classification, the number of subjects for studies averages in the low teens, with single digit subjects being quite frequent. Moreover, some highly influential studies in this area have $N=1$ (e.g., Picard et al., 2001). Also, it is worth noting that the recognition trends reported in the results

section has been observed for each subject in our study, without exception. In other words, the data tell the same story for each of our subjects. While these results problematize most of the common assumptions that underlie research in this area, further, more comprehensive studies are needed to establish whether this is a more general phenomenon.

How to Evaluate the Models

The classical perspective narrates models as representations of emotions that give systems the means to recognize human emotion by classifying sensor input into emotional categories. This move positions models as approximations of emotion: e.g., recognizing fear with $x\%$ accuracy means that x out of 100 times the model correctly identified the emotional category. Here, I consider what models can be said to achieve through the insights that emerged in this chapter.

As we have seen, in order to work, the algorithms require the data — recorded using sensors, in our case — to be partitioned in the classes that we would like the model to differentiate between: fear and non-fear. To arrive at the data and the labeling the algorithms require a number of indeterminacies to be resolved. Perhaps the most important one pertains to the labeling process: picking an emotional category for each data point. As we have seen earlier, emotional categories such as fear are abstractions, i.e., a short-hand way of referring to complex, heterogeneous experiences. In other words, the data points which are used both for training and testing purposes are simple abstractions. Labeling data from the ropes course highlighted the fact that there are different ways of resolving the category indeterminacy. Because of this and other² inher-

²Such as missing values in the sensor recordings. As we have seen earlier in the chapter in

ent indeterminacies, **model testing results such as accuracy cannot be taken as unambiguous information.**

Moreover, from the perspective I take in this study, categorical interpretations of the numerical results are problematic because the way the models are tested is self-referential. This has to do with the way the testing sets have been obtained: in the same way the training sets were derived, i.e., through a particular way of cutting the world — i.e., resolving the category ambiguity — out of many possible. While it makes sense to improve the degree to which the models capture regularities in the data, **the process of evaluating the models must be extended to include the cuts themselves.** This necessarily means examining them in deployment contexts against other ways of cutting the world that will be relevant for the system, e.g., users' ways of making sense of their experiences.

4.4.5 Performing Representations Differently

In this chapter, we have approached emotional experience as a response to and a way of making sense of the world. Sorting out our emotional experiences, whether in a scientific or lay context, involves drawing on a number of resources made available to us — which may include physiological, experiential, cultural, social aspects — in order to arrive at meanings and descriptions of such experiences, i.e., formalisms. To achieve this, we form, relate, and detach aspects of experience. These simplifications and ways of ordering them is how we make our world manageable (Bowker & Star, 1999). In the words of Helen Verran, they effect “a solidity within which to go on to further things” (Verran, 2001, p.

the case of the GSR, to obtain a GSR reading the sensor must send a small current through the body so that it may measure the skin's conductivity. In the absence (or in case of malfunction) of the sensor, the current does not exist and thus the GSR value is indeterminate.

114). As we have seen thus far in the context of human emotion, the practices enacting such achievements are creative: they shape the outcome and there are different ways of doing so, thus effecting different outcomes. Classically, technical approaches to doing emotion have been framed in objective, fixed terms. In this section, I use my experience with building models of emotion to discuss how we may begin to represent and thus do emotion differently using such technical approaches.

Recall from chapter 3 that the traditional worldview assumes an inherent separation between agencies of observation and object of interest. This presumed inherent distance serves as a guarantee that representing the object does not interfere with it and that the outcome of representation stands in one to one correspondence with the represented. In other words, obtaining the correct (i.e., objective) representation requires that the scientist does not intervene in any way that may change the object of study. In this way, what is true (i.e., real) can be thought to guide the act of representing.

The agential realist view I take in this thesis as an alternative metaphysics problematizes the inherent nature of the subject — object separation and instead sees the separation as achieved through practices of engagement. Different practices enact the separation differently and thus may effect different outcomes. As we have seen in chapter 3, the boundary around the object may thus be drawn differently. Moreover, earlier in this chapter, we have seen concretely that practices of representing entail deciding what should count as the object of representation, i.e., (re)drawing the object's boundary. An important aspect that I have highlighted in the previous sections has been that arriving at a representation can be achieved through a number of different steps. These different

courses of action result in different representations; thus, depending on which course of action we take, we do the object represented differently. In contrast to the traditional perspective, representing not only achieves a representation but also the object itself; in other words, our actions contribute to the realities we enact. Instead of being guided by what is given a priori, then, **representing in this framework is guided by an orientation towards the realities we would like to enact.**

Approaching representations differently within this emergent worldview entails identifying alternative ways in which the representation, and thus the represented, may be performed and carefully considering the implications of each alternative. This is at once liberating — now we understand that there is not one fixed way of representing and thus doing the world and are free to choose the alternative appropriate for the current situation — and casts us accountable for the realities we help enact.

Agential realism provides a framework in which these aspects — identifying alternatives and evaluating consequences — may be rigorously examined. Crucial tools are the notion of the phenomenon and the new take on objectivity. Objectivity in this framework requires that we take ourselves as part of the phenomena we are trying to understand. As we have seen in this chapter, this sensibility towards the ways in which we are inextricably involved in the world provides a gateway through which alternative ways of representing the world begin to surface. Acknowledging that we are “meeting the universe half way” (Barad, 2007) compels us to continuously identify the ‘what, how, why, and with what consequences’ of our actions — i.e., how we ourselves participate in these enactments. As we have seen in this chapter nothing may be taken for granted:

no action is *inherently* the correct one. Infusing technical work with a close and continuous examination of the role we ourselves play in shaping the realities we are part of resonates with what Agre calls a critical technical practice. In his words, such a practice “will [...] require a split identity — on the one hand, one foot planted in the craft work of design and the other foot planted in the reflexive work of critique” (Agre, 1997). This work, then, contributes towards a critical technical practice.

The notion of a **phenomenon** makes for a suitable tool to explore critical technical practices:

1. As the ontological unit within agential realism, it already casts us — i.e., agencies of observation — as an integral, indeed constitutive, part of technical practices.
2. It provides concrete practical and theoretical means to extend **technical practices to include the ways in which we engage descriptions** — i.e., representations of technical and scientific practices such as methodologies — **to arrive at the concrete sequence of actions that enacts them on the ground**. To this end, the notion of an apparatus — at once material and conceptual — is a vehicle for exploring how theory and practice co-constitute one another and for tracing other material-conceptual entanglements, more broadly.
3. Representations are the outcome of separations within phenomena. To achieve such separations a material-conceptual apparatus effects a cut within the phenomenon (e.g., the case of recording GSR: a sensor is used to create and separate the GSR from the body.) Importantly, the notion of a cut makes it possible to consider at once what is included and as well as

what is excluded from representing. Thus, together, cuts and apparatuses provide the means to concretely explore how we may do things differently in technical practice: identifying representational alternatives and examining the consequences of choosing one over another as they give us the means to account for what was included and what was excluded.

This latter point reiterates the fact that we are **accountable** for the realities we help enact. A technical practice that is accountable for its doings necessarily implies slowing down so we may adjust our practices to better understand the consequences of our actions. At the same time we must reconsider the rhetoric that accompanies existing technical practices, such as machine learning. For example, discourses around ubicomp — such as context-aware systems (Dey, 2001) and the smart home (Kidd et al., 1999; Park et al., 2003) — and affective computing (Picard, 1997) technologies that engage statistical models emphasize their autonomy: such technologies, we are told, independent of human judgment, may establish the true relevance of lived phenomena such as emotion and human action. This project uncovers the role of the human work that is needed to make the technologies work and thus problematizes the view of autonomy as a defining attribute *inherent* in such technologies. Further, it recasts classification results as inherently ambiguous and thus questions the representationalist usage which rigidly imposes such models on reality. Strategies and discourses around these technologies that fail to take into account the constitutive nature of such performances are not simply setting aside a variable that may be included in the analysis at a later time. Rather, they fail to recognize that these are productive practices and that the aspects excluded through these practices matter. The point is that technical practices cannot continue to ignore such issues without taking responsibility and being accountable for the consti-

tutive effects of their actions. At stake, then, are epistemological, ontological, and ethical issues.

In sum, a critical technical practice that does representation differently focuses attention to our doings, to alternatives, and to consequences. Mapping out alternatives must start with an orientation towards our actions — i.e., the ways in which we help enact certain outcomes and not others — and to the material arrangements through which they are carried out. Finally, choosing an alternative should be guided by the question: which realities do we want to help enact?

4.4.6 Summing Up

This project has taken a detailed look at technical practices of building models of human emotions from physiological data using machine learning methods. The approach we have taken has been to perform all the steps involved in the process of arriving at such models — from the experimental design of data elicitation sessions all the way to building and testing models. An important aspect of our study has been to contrast the work of building models between two different settings: simplified environments such as the laboratory with more realistic — from the perspective of experiences of emotions — environments. Approaching the analysis in this study from a performative perspective, I have focused on the human work that is required to make the algorithms and thus the models themselves work. Uncovering this work has allowed me to rework on the ground — i.e., through practice — a number of aspects relevant to the relationship between represented and representation in the context of human

emotion. Here is a succinct view of the contributions of this project:

- It repositions the classical notion of **uncertainty** — a gap in our knowledge — of the ground truth of experienced emotion as an ambiguity — i.e., **inherent indeterminacy** — in how to categorize emotional experience into clearly delineated emotional categories.
- It highlights the role of **bias** as necessary in order to resolve this indeterminacy in practice, rather than something to be shunned.
- It recasts the process of finding instances of the object of study, fear, as more akin to **curating** (forming, detaching, framing), rather than simply finding the objects as such.
- The process of obtaining concrete instances is inextricably entangled with (re)defining the abstract category the instances belong to (i.e., the ideal category of fear). This is in contrast to the classical view which sees the ideal category as given in advance and fixed. In other words, our practices are **creative**, rather than **descriptive** of reality.
- This realization **opens up new ways of performing representations** as we are no longer confined to the one, a priori correct way of performing representations and makes us **accountable** for the realities we help shape through the representations we perform.
- In the classical framework, objects of the same ideal category (e.g., instances of fear, or even models of fear) are assumed to share the structure of the category. This means that certain aspects are taken to be the same across such instances without proving or empirically verifying them. I examine and problematize some of these assumptions in my study: e.g., that there exist emotion specific patterns in the physiological data.

- The **numerical results of testing the models** — e.g., 85% accuracy — **may not be taken as unambiguous information**. This brings us to the next implication:
- Evaluating models must be extended beyond statistical measures such as accuracy and recall, to include studies in the context in which they will be used in order to better understand how these representations shape the object represented — fear, in our case. The context should go beyond the physical, to include social and cultural aspects together with the specific work the system incorporating such representations hopes to achieve.
- The relationship between theory and practice e.g., methods for driving models of emotions must be understood as itself situated and contingent. In as much as theory can be seen as an abstraction of practice, it does not determine practice, e.g., applying a method on the ground may not be simply reduced to following a number of predetermined steps. Rather, applying a method involves unpacking it and making it meaningful in the context in which it is applied which itself requires dealing with emergent and contingents aspects which may not fully described in the method itself.

CHAPTER 5

MEETING EMOTION HALFWAY: DESIGNING, BUILDING, AND EVALUATING AN AFFECTIVE INTERACTIVE SYSTEM

So far in this dissertation, we have seen that the problem of representation is at the heart of both practical — i.e., every day technical practice — and philosophical — ontological, epistemological, and ethical — concerns. Chapter 3 has shown that representationalism (the assumption underlying the majority of Western scientific and technical efforts, including the vast majority of HCI research, that representations should stand in one to one correspondence with the represented) is entangled in a broader web of metaphysical assumptions which see the world as a collection of entities, given in advance and existing independently of each other. These assumptions have come under much scrutiny in the past decades in a variety of fields and have been challenged by emerging understandings both of social life and physical processes — e.g., experiments in quantum mechanics. In chapter 3, I have outlined an alternative metaphysics, Barad’s agential realism (Barad, 2007), that does not rely on problematic assumptions (such as the *inherent* object subject separation) and instead of seeing the world as given in advance, approaches it as being continuously performed through practices of engagement. Importantly for the topic of my research, agential realism does away with representationalism. In chapter 4, I have examined technical representations through the lens of performativity, studied in the context of representations of human emotions how current approaches — such as machine learning — perform the represented, and proposed ways in which technical practice may be reoriented to represent, and thus perform the world, differently.

This chapter builds on the previous ones and describes the design, construction, and user study of an interactive system performed within a performative worldview. The system I built exploits the technically advanced representations of emotions that were the object and vehicle of investigation in chapter 3: statistical models that classify physiological data from sensors into emotional categories. My intention in building this system has been twofold. On the one hand, I wanted to offer a demonstration of what building, designing, and evaluating from this alternative perspective entails — how all these steps are performed differently, or not, vis a vis traditional approaches. On the other hand, all these steps provide another opportunity to engage further with the issues that are central to this thesis: the problem of representation, the perceived distance between the social and the technical and, ultimately, adjusting our practices so that we may facilitate more compelling human computer interactions.

I begin with an overview of design goals and constraints, following which I describe the construction of a model of fear from physiology to be incorporated in the system. I continue with the design of the system and of the interaction with the system, followed by the technical details. The chapter ends with a user study and a discussion of what it means to evaluate systems within a performative orientation.

5.1 Preliminaries

The system that makes the object of this chapter — Freaky — is an interactive, mobile system engaging sensor-based statistical classification. It is designed to *help its users experience and understand their emotions*. The target emotion cho-

sen for the system is fear. The primary goal of this project has been to create a research system through which to examine how we might build affective interactive systems differently. The emphasis has been on devising a vehicle through which to explore through practice what it means to make and evaluate interactive systems when we no longer rely on representationalism and a priori separations such as the one between the social and the technical. More specifically, we oriented the design of the system and the user study in ways that would allow us to facilitate, experiment with, and understand human-machine performances of emotion.

To show how things might be done differently even when engaging technologies and approaches developed within the representationalist mindset, in the construction of the system I explicitly engage the data sources (e.g., physiology) and the technical methods for deriving representations — machine learning — classically used in affective interactive systems. Yet, I will do so without relying on the representationalist assumption that these data sources and representations necessarily capture the emotion experienced by the user.

Most affective interactive systems are designed for static settings — the user sitting in front of a desktop computer, e.g., playing a computer game. As we have seen in chapter 4, the models' accuracy is much higher in such restricted settings; in more naturalistic environments there is much more ambiguity, which is also reflected in the models' traditional performance metrics, and so it is not clear whether and how to use statistical classifiers in such contexts. We were interested in taking up the challenge of designing technologies that might be useful in such naturalistic settings. Therefore, we envisioned that Freaky would be a mobile system to be carried by its users across a number of outdoors

settings a scenario not unlike that of many ubiquitous computing applications, i.e., the user moving freely which typically includes light physical activity such as walking, hiking, going up stairs, etc.

This decision places some constraints on the design of the system. To afford a comfortable interaction with the system, the weight and the dimensions of the system must be carefully considered, i.e., striving to minimize them as much as possible. The choice of sensors must also take into account pragmatic considerations such as the likelihood of the sensors to remain functional with minimal user supervision and their potential to interfere and to limit the user mobility. Another constraint is related to maximizing battery time which meant we had to take into account how much power the devices require when picking the hardware, e.g., computational platform, sensors, etc. Next, I describe the physiological data we used and the ways in which we built the model of fear to be incorporated in the system.

5.2 Model Building

To construct the model of fear to be incorporated in the system, we drew on the insights that emerged from our study of statistical emotion classifiers from physiological data that was the object of the previous chapter. These insights informed and oriented our efforts to build models for the context of our system. In this section, I describe how we negotiated the circumstances and design constraints specific to Freaky.

5.2.1 Sensors

The sensors used in our previous experiments — heart activity, galvanic skin response, and respiration volume — required wires for the connection to the recording unit. The wires introduced a number of difficulties, including limiting the users' range of movements and data loss due to unplugged wires as a result of user movement. As light physical activity was part of Freaky's envisioned usage scenario, we decided to rely exclusively on wireless sensors. Another aspect when choosing sensors was robustness, e.g., that the sensor wouldn't detach easily during use. During our previous experiments, the GSR sensor proved to be the most prone to such difficulties. We decided to start our investigations with a heart rate sensor and to add more sensors at a later time if needed. The choice was also informed by the fact that features computed based on heart activity data were the most informative in our previous experiments. For robustness reasons, we chose Garmin's Heart Rate Strap, a commercially available heart rate monitor, designed for and used extensively during physical activity.

5.2.2 Data Gathering

The data collection process was primarily shaped by the specifics of our context of design. The data was recorded in outdoors settings similar — some identical — to the ones in which Freaky will be used. The number of people who volunteered for data collection was five ($N=5$), four of whom also participated in the user study which I describe later in the chapter. We informed the participants about the purpose of the data collection and more broadly about what

we were hoping to achieve with it, including details about the system, e.g., its aim to aid user experience, awareness, and reflection on fear. We invited them to explore areas and engage in activities that might evoke experiences that they would consider fearful or related to fear. Our volunteers typically explored places where they had scary experiences in the past from natural areas such as trails and gorges to cemeteries.

We recorded roughly six hours of data from which, together with the volunteers, we curated a number of episodes to be used as fear data, i.e., to orient the models towards the kind of physiological events we would like Freaky to be informed about. For this we had a broad understanding of fear in mind and let our volunteers drive the process of identifying such events. The subjects used a stopwatch for recording timestamps of events that might be related to fear. Here are two representative examples for the kind of events we selected:

- The volunteer was walking through the cemetery late at night accompanied by a friend. While he did not find the cemetery itself frightening, he started talking about intense past experience of fear which made him quite scared. After the recording was over, he wrote the following reflections: *“The high schools stories I described were stories of calling danger’s bluff, or tempting death. There was the story of hitting 110 mph while going over a bridge on a highway, and the story of climbing onto a chimney on a campus building after sneaking onto the roof, almost falling off. Part of what made this scary was that I hadn’t shared those stories with anyone, and even though I had thought about them a lot, and even though I think I’ve changed since then, maybe become less reckless, they still felt eerie to me. Maybe I hadn’t escaped this death drive, only learned to tame it; maybe it could be triggered again; maybe it was*

character building. Other cultures have intense physical trials to mark entry into manhood. Is reckless behavior the release of a primal need to confront mortality that has collided with an obsession with safety: a necessary step in the formation of an adult psyche, something to shatter the innocent sense of invincibility that marks childhood, and perhaps something better regulated by ancient physical rites of passage? I don't know, but the memories of these things that I had done are haunting. It was a (sic) intense fear, but not of something imminent. Eventually, we started talking about something else, and I wasn't afraid anymore. Oh, I should mention that I probably would have been scared much more frequently had I been alone. Talking can be grounding."

- *"When we had started to go up the gorge trail, I only wanted to go a short distance in. There was something reassuring, from where we were standing, about the distance of the waterfall, its framing by vegetation, the angle and distance of the chain barring the path from us. C and I were just chatting, I believe gossiping. I was already a bit freaked out by it all when the path got dark enough to obscure my feet. I also felt that going any closer would be more than I could handle. You couldn't see much in front, just a couple of steps. I was holding C's arm and told him I was freaked out so he got a timestamp. Shortly after, out of nowhere: flashlights in our eyes, manly voices shouting at us. I was so terrified, no idea what was going on, didn't understand who those people are and what they wanted from us. It was as if the sensory information received by my body was being interpreted raw: noises, loud, danger, panic no finer control. They were yelling at us to stay where we were. They started approaching us slowly and asked if we had any weapons in the bags. They identified themselves as the police. I was a little bit more relaxed, although my heart was racing and I was shaking. They wanted ID, etc. and told us we were not supposed to be there. We said we didn't know that.*

When asked about the 5 signs we had passed reading that the trail is closed after sunset, I told them we didn't see them, we were too busy chatting. I kept telling C: "I'm so scared, this is so good!" which probably confused the hell out of the policemen. After interacting with the policemen for a couple of minutes I calmed down. We spent about 5-10 minutes with them but it's hard to tell; time passes differently in situations like that. They wanted to know what we were up to late at night on the trail. I told them we were collecting data and how this is ironic, but they were actually helping out. They were amused (maybe?) and commented "well, if it's for science..." and let us go. I hadn't been that scared in a really long time."

As these episodes demonstrate, we selected intense experiences, though not necessarily homogenous from an experiential perspective: some involved reliving fearful episodes, while in others scary aspects were physically present in the environment. The decision to pick intense experience was motivated by the experimental and analytical insights derived in the previous chapter. Concretely, one of the trade-offs discussed there involves limiting either the physical context or the kind of events used for training in order to align the model's predictions with events of interest in the world. In other words, this decision situated the model by orienting it towards more intense physiological data points and away from lower intensity ones. As discussed in chapter 4, explicitly intervening in how the model is built is valid in a performative framework as the researcher is part of the phenomenon, rather than external to the object of study as is the case in the classic framework.

5.2.3 Model Training and Testing

Out of the approximately six hours of data we labeled about 15 minutes as fear for model training purposes and testing, to which I now turn. Once again, SVMs were the machine learning method of choice. The software library that we used was LIBSVM (Chang & Lin, 2001) as this was easier to work with on the particular hardware platform we chose (which I describe in the section on Technical Implementation). Some of the data features used in chapter 4, such as power spectrogram, can be quite computation-intensive. This introduces a trade-off between computational delays which affect negatively the responsiveness of the system and classification results. To navigate this trade-off, our strategy in choosing features was to start with relatively simple ones and iteratively build and test models and then add more features as needed. The model that eventually became part of the system uses only three features: the normalized heart rate, delta1 (the difference between two successive values of the heart rate), and delta10 (the difference between the current value and the average of the previous 10 values). We have experimented with a number of different parameters for the SVM. Perhaps the most important parameter is the kernel function; it projects the data points into a higher dimensional space in which the SVM may perform a linear separation of the data points into two classes, in our case fear and non-fear. In other words, the kernel function allows for a non-linear classifier. The kernel we have picked based on the classification results was RBF (radial basis function).

We performed 10-fold cross-validation for which I present the averaged results. The model showed good classification accuracy: 97%. This was somewhat expected as the majority of the points had a negative label (no fear). So a more

telling measure was recall: the proportion of fear points correctly identified. The recall in this case was only 27%. In other words, roughly one in four fear points was classified as such. However, throughout our different experiments with different features and parameters, while we kept an eye on the traditional metrics such as accuracy and recall, a different measure proved to be more important from the perspective of what we were trying to achieve with Freaky. Our aim for Freaky was not necessarily to detect each data point (corresponding to a time window of one second), but rather to identify a relevant event which typically lasts much longer than one second. This level of granularity, then, was adequate for our purposes. In other words, we modified the evaluation criteria to suit the specifics of the system. From the perspective of this metric, the model obtained using the features and parameters described above identified all the relevant events (however, part of this performance might be attributed to over fitting). Consequently, we decided that the first prototype would incorporate this model and if necessary we could return to further experimentation if the model's behavior in use called for it.

5.3 Meeting Emotion Halfway: Designing Freaky

The classical perspective sees the social and the technical as a priori separate and distinct entities. As such, in the case of emotion, it would appear that there is an *inherent* gap between technical approaches to human emotions, such as the models, and social understandings of human emotion. Concretely, using machine learning on data from only one sensor and computing only a small number of features would result in a significant rift between the model's understanding — a binary prediction — and users' emotional state. Typically, systems

in affective or ubiquitous computing that use statistical models approach their use from a representationalist perspective. Such systems rely overwhelmingly on the accuracy of the classification. From the classical perspective, the gap must be narrowed by incorporating more technical elements — sensors, features, more comprehensive learning strategies.

In a worldview that treats realities as emergent, the social and the technical co-articulate one another and emerge as a result of particular enactments, rather than being given in advance. Differences and similarities between the two, then, are continuously in flux. It follows that understanding how humans and machines do emotion differently is a first step towards figuring out how they may be put in a constructive dialogue. Statistical models of emotion give a particular perspective into human emotional experiences. Specifically, they cut emotion based on physiological features that resemble those of the instances used during training. People's practices around emotion enact different cuts. Designing a system from an emergent perspective means providing a way for the machine cut to inform and thus participate in users' cuts. In Freaky's case, this would make it possible for users to engage the information provided by the model when making sense of their experiences. To understand how this might be done, we must take into account similarities and differences between the ways people and machines cut emotion.

A significant part of chapter 4 has explored how techno-science does emotion differently than humans. Both human and machine practices are heterogeneous and continuously evolving and responding to one another and to other practices. Therefore, the differences between the two cannot be fully described, nor can they be assumed to remain fixed. However, it is possible to identify

which dimensions are emphasized by techno-scientific practices and which dimensions typically relevant for humans are excluded or marginalized by such practices.

No doubt, techno-scientific emotion cutting practices are different from the ways people perform emotion in their lives. While it is easy to focus on their shortcomings (e.g., much of the richness and diversity of embodied experience seems to be lost) machine cuts of emotion such as the statistical models described before which may be readily embedded within computational devices are *at once less and more* than lived emotion. Less is obvious: this has to do with lived emotion's cultural and social embodiment (e.g., Boehner et al., 2007) and with its multiplicity — these aspects tend to disappear from view in machine interpretations of emotion. More in the sense that material resources that may not be available in interactions that are not technically mediated — such as patterns within physiological data — now become available. Thus, as emotion may now inhabit a new medium — digital — and may be located in computer bodies, not just human, it is significantly reshaped along cultural, social, physiological, and statistical dimensions.

This suggests that a way to bring the machine's perspective into the users' world is to orient the interaction with the system in such ways as to provide an embodied way for the user to engage the machine's perspective. By embodied I do not mean only physical, but also socio-cultural embodiment (Dourish, 2001). How might we achieve that when designing a system? My answer to this question was informed by Helen Verran's analysis of how abstract objects — such as numbering systems — that emerge through different practices are made to work together by creatively linking them in performance; in the case

of the Yoruba classrooms that were the site of Verran observations, the linking was achieved through a ritual (Verran, 2001). As such, we conceptualized the interaction with Freaky as a ritual (a sequence of actions to be repeated at certain times) through which Freaky's perspectives into emotion may be creatively connected by users to their own.

Before going into the specifics of the interaction with Freaky, I describe its conceptual and physical design. Inspired by Romero et al.'s alien presence approach (Romero et al., 2008), i.e., framing the system as having an idiosyncratic point of view, we conceptualized the system as an artificial companion with its own understanding of fear. We did so as a way to frame in design the insight that the computer's representations are not the same as users' experiences. To this end, Freaky is endowed through interaction design with its own fear and fear associated behaviors. For example, when Freaky classifies the user's physiological data as fear it becomes scared; it manifests it by vibrating and intensifying the audio output. Freaky may thus be thought of as experiencing parts of the users' experiences through human-machine emotional contagion. To further orient the interaction towards exploring similarities as well as discrepancies between the users experience and the system's perspectives, Freaky's physical design features a familiar and, at the same time, unusual shape and surface (figure 5.1). These physical characteristics are intended to give the users a sense that Freaky might have peculiar, yet hopefully useful, insights into human emotion.

We hoped that endowing Freaky with its own fear and associated behaviors would allow the user to create a narrative about the machine's fear and understandings of fear as its own, although linked to the user's physiology. Through both the overlaps as well as the differences between machine and user emotion,



Figure 5.1: Freaky's shell.

the user would have the opportunity to reflect on her own, by continuously being prompted to sort her experiences as well as Freaky's.



Figure 5.2: Interacting with Freaky.

We envisioned the interaction with Freaky in the following way (illustrated in figure 5.2):

- in “normal” mode, Freaky is strapped to the user's chest. It produces continuous audio output generated in real time based on the user's heart rate. The person thus has the opportunity to link changes in the sounds to her

body's reactions. Similar to the way the Tableau Machine generates visualizations of the activity in the home from sensor information (Romero et al., 2008), the mixing function we used to generate the sound file is nonlinear, to avoid repetitions which may lead to user disengagement.

- Freaky goes into “freak out” mode when the model of emotion predicts the label “fear.” The system begins to act scared: it starts vibrating and the audio output intensifies.
- “calming down” mode: to return to “normal,” Freaky requires attention from the user: the system checks that data from the pressure, light, temperature, and acceleration sensors fall in a soothing range, i.e., avoiding extreme values. The sensed data is likely to fall within the soothing intervals when the user is petting, protecting, holding, or rocking Freaky like a baby. We specifically chose these user actions because they are likely to also calm down the user. Our decision was inspired by the strategy to pull the user into an affective loop by choosing an input gesture for interaction that may further induce that emotion (Höök, 2008); here, however, we reversed this strategy to invite the user to calm down if scared. Calming down Freaky can be done through a combination of sensors; once again, the process is not linear, inviting the user to explore different ways of achieving this. Put simply, the more sensory inputs are engaged in calming the system down, the quicker it returns to normal. The intensity of the vibrations is inverse proportional to the calmness of the system — at maximum level while in “freak out” mode, decreasing gradually through calming down motions. If the model classifies the current heart rate data as fear when in calm down the system returns to freak out mode.

We were concerned that in a potentially serious situation, the user might

be distracted by the system. Our discomfort was alleviated by a recent study showing that when very stressed out, users tend to ignore real time physiological feedback, thus the feedback does not increase the burden on the user (Ferreira et al., 2008).

5.4 Hardware Implementation

After a great deal of experimenting, to build a prototype of the system I have used:

- a BeagleBoard unit, a low-power, low-cost single-board computer, as the main computational platform is. The unit measures only 75 by 75 mm and provides all the functionality of a basic computer.
- an Arduino microcontroller to handle the sensors and the vibrator motor: it relays sensor values to the BeagleBoard and BeagleBoard commands to the vibration motor.
- An accelerometer and light, pressure, and temperature sensors.
- a vibration motor.
- six 1.2V rechargeable batteries.
- A current converter to 5V DC.

5.5 Software Implementation

Computer programs on the Arduino and the BeagleBoard make up Freaky's software body. A simple program¹ on the Arduino communicates through a serial port with the main computational platform, the BeagleBoard unit. It receives and executes commands to read sensor values and update the power level on the vibration motor, thus changing the intensity of vibrations. On the BeagleBoard two programs ran concurrently are responsible for most of Freaky's behaviors. They are implemented in Python. *Freaky.py* sets up the wireless communication link with the Heart Rate monitor; receives heart rate values every second; sets up the communication channel with the Arduino; sends read sensor and update motor commands to the Arduino; computes features based on the heart rate data; runs the fear classifier, i.e., the model; and, implements Freaky's behaviors: freaking out, calming down, normal mode. The other program, *SoundServer.py* receives the heart rate value from *Freaky.py* every second and uses it to generate Freaky's audio output.

5.6 Studying Freaky in Action

In this section, I describe an exploratory user study of Freaky and the insights that emerged from it. I approached this study from a performative mindset. Its aim was twofold:

- to examine performances of emotion that emerge in interaction and to understand how Freaky participates in them

¹Arduino programs are written in C/C++.

- to understand how we may begin to think differently about studies of technology in use and evaluating interactive systems by contrasting it to a classic evaluation



Figure 5.3: One of our participants and Freaky.

The participants were self-selected — in the sense that they explicitly manifested interest in exploring their emotions using an interactive system — from my social circles. This orientation towards my social circles was motivated as before: by continuing to defamiliarize the view of the researcher as an outside, detached observer. A total of eight people participated in the study, including myself, of which five were women. Participants' ages ranged between 24 and 37. Four of them had also participated in the data collection sessions and thus their data was used to build the classification model incorporated in the prototype.

The study focused on the participants' interactions with Freaky. Prior to using the system, I met individually with each participant and explained how Freaky was built and how they might interact with it. This included a description of the model mapping physiology to fear, specifically detailing how it was constructed and that the model's classification is based on statistical similarities between users' physiological data and the points that were used to train the model. Each participant had the freedom to interact with Freaky in a way that suited him/her: e.g., choosing how many times, where, when, with whom, and how long to use the system. In some cases, I accompanied the user while s/he was using Freaky and audio recorded the use session. After using the system, I interviewed the volunteers in order to debrief them on their experiences. In the days and weeks that followed their interactions, some of the participants sent me via email or expressed in person further reflections on and insights gained from their interactions with Freaky. These interactions were also documented as part of the study.

Table 5.1 presents an overview of the user study data set (a star next to the user's initial indicates whether Freaky freaked out during the interaction with that user):

User	Duration	Context of Use
C	2 hrs	Night; accompanied; gorge trail, town, campus
M*	1 hr	Night; accompanied; cemetery, residential area
U*	1 hr	Day; accompanied; residential area
H*	3 hrs	Day and night; alone; indoors, packing, watching a movie
Q*	3 hrs	Day; alone; forest walk, house-sitting
K*	1 hr	Day; indoors; rehearsing anxiety provoking talk
S*	1 hr	Day; alone; at, to and from doctor's office
L*	5 hrs	Daytime; alone; indoors and outdoors

Table 5.1: Overview of the users study.

In the following, I describe three episodes that are suggestive of the kinds of interactions afforded by the system and the insights that emerged from the study. I have picked these examples because they foreground the issues that are central to my research: examining performances of emotion through human-machine interactions and how Freaky participates in them, which I will discuss after presenting the examples.

5.6.1 Max and the cemetery

Max decided that a good place to explore with Freaky was the cemetery at night time. He finds cemeteries to be extremely frightening as he is afraid of ghosts. We met at his apartment. Max was very tense pretty much as soon as we left his apartment. One could say that he was afraid of his own shadow. He adopted a “fight or flight” position: tense, constantly looking around, a bit paranoid, fists clenched around the LED wands he had brought along, arms resting on top of the baby carrier; it was as if the wands were some sort of weapon. He commented repeatedly on the music generated by Freaky being very spooky, not unlike Halloween haunted house style. He mentioned he was too tense to pay attention to nuances or changes in the music. We walked through the cemetery for about 20 minutes or so. Even though he mentioned being extremely scared a number of times, Freaky did not freak out.

After leaving the cemetery we were discussing possible explanations as to why Freaky did not detect fear while in the cemetery. Max noted that he could not point out a specific fearful event, but rather lots of tension which did not really culminate in any way. He commented that maybe the reason Freaky did

not pick up anything in the cemetery had to do with the fact that the “fear wasn’t real.” I asked him to clarify: he meant that he is scared of ghosts, which are not real entities.

We then started talking of “real” fear — the kind that actually materialized in some way or another, rather than just the possibility of fear. Max told me that the scariest he had ever been was about seven months before. As soon as he started talking about that event, Freaky started vibrating:

“[...] the night I found C. E and J said that... [freaky goes off, he pauses] yeah... reliving that moment. [starts tending to Freaky] E and J said they’ve never heard me scream so loud. [changes the topic, speaks about how people would find Freaky enjoyable because it vibrates. In the meantime, Freaky continues to vibrate. Max eventually returns to what had happened that night]

He was hanging out with three close friends, late at night in one of their apartments. As they were about to fall asleep, Max heard a noise in the kitchen, like something dropping on the floor. He rushed to the kitchen and found one of his friends, C, on the floor, lying in the pool of blood, and holding a knife in her hand. She had big wounds in the neck, chest, and legs.

“I haven’t realized just how scared I was when I found C. There was just too much going on through my mind... so confused She was lying in a puddle of blood, I grabbed the knife away from her, called 911, got towels and put pressure on her neck wounds, and once the cops got there I was talking to the cops while still putting pressure on her wounds. [...]”

Max explained that at that time he was simply focusing on what can be done. Once his friend was taken away by an ambulance, he was in shock which lasted

for days, maybe even a couple weeks. Eventually he began to grasp the trauma he had gone through and started integrating this experience in his life. Now when he thinks about that night he finds it scary:

"I guess I do have some issues um not issues, but it still scares me. It was intense. [talks about how Freaky didn't pick up anything in the cemetery]. well, that was different. that was like a real moment, I was scared for someone else's life... earlier I was just scared for ghost stuff. [...] All I cared about was C. oh, why would she do that? I often ask myself "why?" [talks for about 5 minutes about the reasons that may have contributed to his friend killing herself]"

5.6.2 Uma and Tyler

Uma had just started wearing Freaky when she received a call from her friend Tyler. He was driving by the place where she was staying and wanted to say goodbye. Uma and Tyler had met approximately a week before. She was visiting from Europe and they happened to be part of the same group of friends attending a music festival. They had been getting along famously, but it was time to say goodbye as she was heading back to Europe the following day.

Tyler parked his car across the street from where she was standing. He crossed the street and greeted her. Before she could reply, Freaky started to vibrate vigorously. Tyler seemed confused as to what was going on. Uma picked up Freaky from the chest carrier and started attending to it. At the same time, she explained that she was part of a study and told him a bit about how Freaky works. He was amused and suggested she was just too excited to see him and had to calm herself down.

Was she excited to see Tyler? In the discussion that followed after her usage session, she explained: *"Not more than usual. I remember thinking when I saw him that I wouldn't want it to go off right then. But it did. And then I got really excited thinking that it picked up the emotional substrate of that encounter: I was exposed emotionally before saying anything. It [Freaky] participated in the encounter. But it wasn't clear what it was [picking up on]. Probably excitement"*

Does she think Freaky was on to something? *"YES. Something I wasn't aware of. I loved it. Maybe it [her not being aware of her reactions] had to do with me being absorbed by meeting Tyler and not fully attending to my emotional reactions. It made me aware that something is different emotionally, which may have otherwise gone unnoticed."*

What was it like calming Freaky down? *"It vibrated pretty hard for a few minutes and so I had to insist with petting it. Gradually, the intensity decreased. I was surprised it took a while, just like a little person."*

After visiting with Tyler for a while, it was time to say goodbye. Things got a bit emotional as they were not sure whether they will see each other again. As Uma was saying goodbye, she cried out: *"why isn't it freaking out now?"* to which Tyler replied: *"because now you're sad."* Later, during the interview she explained: *"Freaky didn't react. It's a mystery I don't know maybe it doesn't pick up on low emotions."*

5.6.3 Writing Anxieties

The third episode describes one of my interactions with Freaky. I was at home one morning working on chapter 4 of this thesis. Writing wasn't going very well. I was experiencing a lot of anxiety about my dissertation: I was a bit behind schedule and was nervous about finishing on time. I picked up Freaky and turned it on. Just holding and petting it as I was pacing around my apartment was great. It felt very soothing, although anxiety was still there, but I felt that I was keeping it under control. It felt as if I was not alone: it was like being with myself, but mediated: a part of self that was externalized, materialized in the world, which allowed me to at once express it and take care of it, in order to attend to it more fully.

After a while I sat down again and picked up where I had left off all the while holding Freaky. After another 15 minutes of work it freaked out. I pulled it out of the sling, began walking around the kitchen tending to it. I was quite surprised how long it took to begin calming down, at least 5 minutes. I noticed some very interesting echoes in the music something I had never heard before which was quite surprising given the amount of time I had spent listening to Freaky's audio output — using it, but mostly building and debugging it. It made me think how that particular situation might be different than what it had 'seen' before. That particular time period was busy not just with writing but also having to vacate my apartment, leaving Ithaca in a couple of months, making various travel, working, and living arrangements for the months to follow. There was a mix of sadness as this period marked the end of era — it felt as if something was dying inside of me and it needed attention and expression — and the excitement of new beginnings. It also made me think how when

working I tend to get so absorbed in what I am doing that feelings and subtle sensations in my body — like getting tensed up — pass unnoticed. When they finally register consciously, it is often after I get restless and all kinds of anxious.

In the meantime, the intensity with which Freaky was vibrating had been decreasing but very, very slowly. It never fully calmed down: there remained a subtle vibration and buzzing noise. It made me think that Freaky was right on: the anxiety and restlessness was definitely still there, but not quite as strong as in the beginning. After a couple of minutes the system began a sequence of consecutive reboots, a sign that the battery level was low. I then realized that this was responsible for the continuous but low vibrating mode: when the power levels are low, the main computational platform is turned off, but the Arduino controller remains on. This means the motor vibration levels could not have been changed — as a program running on the main platform would be the one issuing that command — and so the motor remained in a low vibrating mode.

5.6.4 Intermezzo: Prefacing an Agential Realist Take on Realist Evaluation

Before continuing with the themes and insights that emerged from the user study, I want to first discuss how an agential realist take on evaluation may be different from what may be typically expected in traditional forms of evaluation.

Broadly speaking, evaluation is concerned with validating technological so-

lutions. In other words, evaluation aims to answer the question: **does the proposed technological solution work?** Evaluations in the classical framework seek to do so by measuring the effect caused by that the proposed solution. For example, Wayang is an affect-aware tutor that teaches math and prepares students for standardized tests (Arroyo et al., 2007). The system manipulates a variety of sensor inputs to infer students' interest and adapt the interaction accordingly. To evaluate the effectiveness of their tutor, the researchers who built Wayang compared the pre- and post-usage results of 3 groups: a no-tutor control group, a group using the tutor without the affect module, and a group testing the tutor with interventions driven by the system's recognition of student emotion. The study reported statistically significant improvements in the third groups' scores and passing rates on standardized tests. This meant that a causal relationship between the tutor's interventions and students' improved performances has been shown, allowing the authors to conclude that their technological intervention was successful.

A crucial distinction between agential realism and the traditional worldview pertains to the separation between the object of study and agencies of observation, as I have discussed at length in chapter 3. The latter perspective assumes the separation to be inherent in the world, while the former shows separations to be performed within phenomena and open to different enactments. The nature of separation is relevant for matters of system evaluations because in the classical worldview the object of study — the system — is assumed to be inherently separate from the rest of the world (such as, its context, users, etc.), while agential realism sees this separation as achieved through a series of practices (which, for example in Freaky's case included model building, system construction, interaction design, and the interactions with the users).

These differences between the two perspectives have consequences on matters of causality. The agential realist view sees cause and effect entities as emerging in the phenomenon and thus they are ontologically dependent on the phenomenon and the specific apparatus used to effect the cut. Cause and effect entities (such as system and measurements of behavioral change), then, may be separated differently, for example by using a different material-conceptual apparatus to perform the separation. Consequently, it may not be assumed that a causal relationship between entities carries to other phenomena, as other phenomena may cut differently the boundary between the entities. The conclusion that emerges is that causal relationships may not be assumed to be universally valid; rather, they are dependent on the particular ways in which the world was cut in the phenomena observed. Put it differently, the knowledge that emerges must have the form: “if we separate the world in this way, this causal relationship may be observed.”

Concrete examples will help further clarify these differences and what is at stake. Built within the traditional framework, the affective tutor described above is inextricably entangled with the assumptions of that particular worldview. These include inherent separations between entities such as system, users, context, as well as an understanding of emotions as clearly defined events in the world which may therefore be treated as another type of information (see chapter 4). But agential realism problematizes the inherent nature of these assumptions and shows them to be performed. From that perspective, the system, through its construction and framing by the researchers, enacts the entities enumerated above as separate and performs emotion as a type of information. Further, through experimental design, evaluation is set up so that it too includes the assumptions built in the system and contributes to their enactment. Concretely,

the specifics of the evaluation, such as dividing the subjects into three groups of which two are control groups, performs the separation of the tutor from the rest of world, in such ways as to construct it as the cause entity for which effects may be tracked.

This does not mean that the agential realist perspective sees the result of the traditional evaluation as invalid. Rather, it contextualizes this result: while the classical framework understands the causal relationship to exist independent of the framework itself, through the lens of agential realism we understand that this relationship is an achievement, and a powerful achievement at that: through certain practices that enact the metaphysical assumptions, the causal relationship may be achieved, i.e., it materializes and becomes real. In other words, the broader context of the metaphysical assumptions the system enacts exposes the causal relationship as achieved through particular ways of understanding and doing the world (i.e., as a collection of separate entities), rather than uncovering a relationship that occurred in the world independent of our practices. Put it differently, rather than uncovering a truth inherent in the world, evaluation validates one way of doing reality that achieves a particular outcome.

An agential realist lens, then, makes visible that other metaphysical positions exist and that our practices implicitly enact them not just through system design and construction, but also when it comes to system evaluation. Together, all these enactments shape the kind of outcomes that may emerge from practices of evaluation. For instance, in the classical framework system building practices enact entities (such as system, user, context) and relations between them as *fixed* and *given in advance* (i.e., existing independent of practices of investigation).

It shouldn't come as a surprise, then, that a classical evaluation figures fixed entities in fixed relations e.g., the technological solution of including affective information in the tutor, on the one hand, and users' test results, on the other, as causally related. In an agential realist framework, the situation looks different because this worldview figures entities and relationships as continuously *in flux* and *emerging* through specific enactments. Thus, performed within an agential realist framework, Freaky's evaluation is different than a classical evaluation in that it features performances of various entities, such as emotions, subjects, and relations in flux. This is an expected consequence, if perhaps disconcerting at first for readers accustomed to traditional evaluations, of adopting a different metaphysical position: agential realism.

In what's to come, I use the three examples before together with other observations from Freaky's user study as vehicles to:

1. point out changes in the nature of observations emerging from evaluation due to the different worldview adopt in this dissertation: e.g., focus on performative aspects, emergent attributes in flux (e.g., accountability and autonomy)
2. examine how specific design decisions have contributed to the interaction with the system, i.e., what we may learn from our experiences with Freaky that might be useful in future designs

I return at the end of the chapter with further meta-observations on practices of evaluation.

5.6.5 Understanding Human-Machine Performances of Emotion

Our design aim in building Freaky was to aid users in expressing and making sense of their emotions and experiences. It should not come as a surprise that the overarching theme that emerged from people's encounters with Freaky and the discussions that followed was that of people sorting themselves out through interactions with the system. The three episodes presented before show emotion, self and other, and relationships as continuously performed. In Max's case, reliving an event with and through Freaky contributed to his ongoing process of making sense and of and reconciling previous traumatic experiences. Further, the writing episode recounted earlier foreground my efforts to understand my anxieties and work with them. Further yet, Uma's experience using Freaky speaks of the system's behavior prompting her to explain its workings to Tyler and in the processes working out her emotional reactions. This entailed performing various emotional interpretations, together with Tyler, and in the process performing their relationship: reaffirming and strengthening their connection. That episode highlights how emotion is inextricably entangled with people's actions and broader aspects of people's lives: here, emotion contributes to performing Uma and Tyler's relationship and, at the same time, the relationship shapes their emotional experiences. The picture that emerges from these episodes is that of people sorting themselves out, to themselves and to others, and in the process performing themselves in certain ways and not others. What I want to point out here is the shift in focus brought about by the agential realist framework towards how emotion is performed and its role in the broader context of people's lives and away from the classical perspective of seeing emotion

as an a priori separate entity, a self-contained phenomenon.

Our design strategy to achieve the system's goal of aiding users' emotional expression and understanding was to design a space for machine and user perspectives to be *linked in performance* so that resonant and dissonant aspects between the two may be worked out. We implemented this strategy by designing the interaction with the system as a *ritual* connecting user and machine. For this, we drew on the notion of *embodiment* (e.g., Dourish, 2004; Klemmer et al., 2006). As expected, observations pertaining to these design strategies and approaches surfaced in participants' comments and discussions. In the following discussion of the user study, I begin by focusing on these strategies. This will allow me to (1) understand the ways in which they worked and didn't work, (2) explore further consequences of engaging them, and (3) derive further insights pertinent to designing interaction from a performative perspective.

I begin with observations concerning Freaky's embodiment. Most of our participants pointed more or less directly towards Freaky's physical embodiment as a grounding factor that allowed them to work through their experiences. One of the participants, Q, used Freaky during her stay in the unfamiliar house she was house-sitting for a few days. She commented that when she would be scared by eerie noises in the house, Freaky's vibrations and her holding and petting its body helped her "*get out of [her] head.*" She would then focus on what in the house could have caused the noises and then did something about it, e.g., confine the dog to the downstairs area. This comment speaks not just to Freaky's physical embodiment, but also to how socio-cultural aspects pertaining to Freaky's design, such as its behaviors and the modalities through which the user interacts with it, are always already suggested by its physical design,

as explicitly articulated by C:

"There are ways in which we want to please the machine too. We wouldn't let it to keep on buzzing. I wonder if this is because I know it is supposed to get some soothing movements. This is already a language, chosen for a reason; just like purring. It's not like you get a text message "you're scared right now." So I wonder then, an important aspect is that the device has a body. I also recognize the arbitrariness of the body. You could have skins over it like an ipod customize Freaky For me it fits. There's a natural place for it, to introduce it in the world. [] It can blend in for the most part. Then, it acts as a surface on which to project our needs, fears, aspirations, rewards. It almost feels like therapy: you work with mind objects, but in an embodied kind of way Isn't that like transference? Is there a word for transference to machines?"

Thus, a first take-away point emerging from Freaky's user study pertains to the particular ways in which Freaky's embodiment mattered. Freaky's physical and socio-cultural body played an important role in facilitating its participation in the world. Such a body allowed Freaky to fit in the world or, more to the point, gave Freaky's users the means to fit Freaky in their world. To this end, C described his time spent with Freaky and a friend as a preview of *"the family of the future: two gays and a robot baby,"* or what we might call a post-natural ecology.

These observations point to specific physical and socio-cultural design features that allowed users to seamlessly weave them together in interaction, and through them to navigate between themselves and the system, between their emotion and what the system might be reacting to. H's comments speak to this point: *"I think the shape of Freaky lends itself well to wanting to care for Freaky, it's easy to hug him and tell him everything will be ok."* This back and forth achieves

a place for Freaky in users' world and in their ways of doing things. From a performative point of view, the material (i.e., physical) and the conceptual (e.g., socio-cultural) are inextricably entangled. Therefore, supporting their co-articulation through system design is likely to contribute to useful interactions. As seen here, this may be achieved by designing a coherent system from the perspective of its embodiment: one in which the physical and the socio-technical resonate and build on one another.

Shifting Accountabilities

In all the instances of use presented here, and indeed all the use sessions, the very presence of the device changed the way people made sense of their actions and reactions. This is an expected consequence when working within the agential realist framework and points to the constitutive nature of the role played by Freaky. Here, I detail how matters of accountability materialized in practice so that I may identify how design features contributed in this arena.

The scenario in which aspects pertaining to accountability are most obvious are those of the device 'freaking out,' i.e., predicting fear. As we have seen, this provokes those around the system into explaining what might have caused this behavior. Alternatively, cases in which the system behaves in a more passive fashion, i.e., not vibrating, such as during Max's cemetery walk and Uma parting ways with Tyler, may themselves invite users' explanations. Importantly, these cases show that more than simply demanding an explanation, Freaky contributes to the **formation of an event** in the flow of participants' experiences, e.g., the beginning of the Uma and Tyler's encounter. Firstly, this constitutes a concrete example of how Freaky enacts certain realities. Secondly, it is directly

related to our decision to conceptualize the system's behavior as a set of clearly delineated states that are visible — or more accurately, audible and palpable — to the users and those around the system.

Further, the system offers new resources to be worked into people's explanations. For instance, unusual echoes in the sounds generated by Freaky made me consider how that particular situation was itself unusual. What is remarkable here is that the system managed to surprise me, i.e., someone intimately familiar with it. This had to do with the non-deterministic way in which we programmed the sound generation. While the user heart-rate is taken into account, e.g., by adjusting the frequency and amplitude of the sounds produced, the process is not fully determined by the heart-rate as the overall audio effect is achieved by combining two sound channels. As such, even after long term usage, users may still find surprising sounds and echoes which may be worked into their explanations. These observations point to non-deterministic, generative features of system behavior as important aspects in users' differential ways of accounting for themselves.

Other examples from the usage sessions allow further insight into how Freaky contributed in the changing landscape of accountabilities. These instances highlight Freaky playing a more active role: provoking — or, at a minimum, contributing to — fear. One such instance is Max's walk through the cemetery. He pointed out repeatedly that he finds the system's audio output to be quite scary. While this observation must be understood in that evening's broader context: the cemetery, darkness, Max being afraid of ghosts, etc., it nonetheless points to Freaky contributing to Max's experiences of fear. Another participant remarked a different aspect of the system that might itself cause un-

easy feelings:

“you’re wearing a ridiculous device in public and you can’t pretend that it’s invisible. The heart rate strap is invisible, but the rest isn’t. [...] This might lead to [user’s] paranoia of the device. The obvious social nonconformity of toting this thing around means that, one, either you’re comfortable with it or, two, you just have to own it When we weren’t with people we were in scary situations otherwise, we were around people, different kind of fear, paranoia. [...] But, two people make a place for Freaky as the baby; otherwise, it would probably stand out more. I’d feel vulnerable with it, alone maybe at least awkward, not casual... It’s context dependent. It’s easier doing that here [in Ithaca], not a big deal [excerpt from C’s interview]

That Freaky’s audio output may be scary and its very presence anxiety provoking in certain situations foreground ways in which the system itself may be held accountable for people’s reactions. These instances speak of Freaky helping enact certain realities and not others, i.e., that the system plays a constitutive role in the world which is, as noted before, an expected consequence with the agential realist framework. Moreover, they point to strategies in which designers may draw attention to the constitutive aspects of interactive technologies by disrupting user’s expectations. One example is the inclusion of design features which may themselves cause unanticipated consequences for which the user would find the system itself accountable.

In conclusion, the use instances in this section highlight (1) changes brought about by the agential realist framework, e.g., the field of accountabilities as continuously in flux and always situated, i.e., entangled with the specifics of the situation of use, and (2) design strategies through which the system may specifically intervene in that field. Furthermore, they speak of the concrete, material

ways in which Freaky contributes to the enactments of different realities.

On Autonomy

Another aspect of interactive systems that must be approached in a performative framework as continuously performed in interaction is system autonomy. As expected, an important opportunity for analyzing how the system contributed to these enactments was our participants' reactions to Freaky's fear classifications and the behaviors through which those were manifested in interaction. These sources highlight that, while users perceived the system as a separate entity, the relationship with the device enacted by them was often indicative of the degree to which they were able to relate the system's perspectives to theirs. When the perspectives were seen to resonate with one another, or at a minimum, were temporally synchronized, Freaky would be figured as a separate, yet closely related, dependent entity, as suggested by Uma's comment that Freaky acted *"like a baby picking up mom's intense emotions"*.

In contrast, when users worked out a disconnection between their perspectives and the system's reactions, they projected more autonomy onto Freaky figuring it as not only separate, but also independent:

"I was surprised that Freakster didn't react during the time I was packing, I hate packing and was also under time pressure, so I was sure Freaky would be nerve too. Alas, he was snug as a bug in a rug in the baby bjorn. [...] Surprisingly, as I write this and reflect on our time together, I realize that I don't see Freaky as a manifestation of me or my emotions/feelings, but as a separate being ... a reactive sentient-like being that liked being held and calmed down." [excerpt from H's email detailing her use

trials]

Perhaps surprisingly, the picture that emerges is that users related to the system more and thus perceived it to be more useful when they figured it as being less autonomous. As for design implications for supporting users' continuous and differential figuring of machine autonomy, our design strategy to support the dynamic linking of human and machine perspectives by designing the interaction with Freaky as a ritual showed promising results.

On the Nature of the Gap

It is worthwhile pausing here to mention that from the classical perspective the interactions between Freaky and the participants described in this section could be seen as revealing a socio-technical gap with respect to human emotion. Indeed, if we separate users and system and focus on how each side contributes to performances of emotion, there appears to be a large distance between the two kinds of perspectives. Yet, such an analysis does not do justice to the dynamic nature of these perspectives and, indeed, their entanglement: clearly separating the two is not possible at the very least because users account themselves differently in the presence of the device. Moreover, examining the situation in terms of human-machine performances of emotions reveals difference and sameness between machine and human perspectives as performed and contingent. Thus, many, temporary gaps may appear between perspectives rather than the more or less static disjuncture figured by the traditional perspective. These gaps, or differences, are continuously in flux because the relationship between human and machine perspectives continuously co-articulate one another and are re-worked in interaction, as detailed above. We begin to see that from an emergent

perspective, these differences, or gaps, may be thought as generative as they may engender new understandings and, importantly, changes in perspective, thus constantly reworking differences. These gaps, then, are not something to be done away with, rather they need to be exposed so that they may be examined, engaged, and transformed. Put it differently, difference is where the action is (Dourish, 2001). Dourish has suggested a similarly generative view of the gap (Dourish, 2006; Rode, 2011); here, we get a concrete sense of why this is the case, how it ‘happens,’ and why it matters. I will return to discuss the gap(s) from a performative perspective in the conclusion chapter.

Positioning Freaky

The user study also affords a close examination of how the model of emotion that Freaky incorporates factors in performances of emotion in interaction. The examples presented here suggest that Freaky’s interventions oriented users’ experiences and reflections towards fear. This may not be surprising given that I had told users that Freaky leverages a model of fear. Users’ experiences and interpretations around fear were quite rich pointing to a wide range of experiences around fear: existential fear, fear of ghosts, fear for someone’s life, fear of an imminent physical danger such as slipping off an icy trail into the gorge, and so on. This points to Freaky’s participation in a set of heterogeneous experiences, which may come together under an abstract, yet broad, category of fear.

Perhaps surprisingly, the systems’ orientation towards fear did not preclude other high energy emotions from being brought into view, such as Uma’s excitement. Interestingly, sorting out her experience included questioning whether

the system's response — the model's fear prediction — was random (this fact emerged in subsequent discussions about that event). That her experience was certainly not related to fear made her wonder whether Freaky randomly freaked out; she concluded that this couldn't have been the case, as it would have been too great of a coincidence as that particular time was the only emotional event during the one hour she had used the system. Another user, S, challenged the system's response as pertinent to her emotions: she dismissed Freaky's buzzing as having anything to do with her emotion, and instead attributed it to exertion from walking up a steep hill. These examples show that in spite of being actively involved in shaping the emotional and interpretational landscape in certain ways, Freaky does not dominate the user by imposing its perspectives on the users'.

To date, systems using statistical models of emotion have been built exclusively within the representationalist worldview. Consequently, such systems substitute the interpretative work normally performed by humans with the model's prediction, as the latter is thought to stand in one to one correspondence with users' emotion. Furthermore, the models' predictions are typically invisible to users, thus casting users in a passive role. Implicitly then, the machine's perspectives on human emotion are rigidly imposed on reality. In contrast, as we have seen in chapter 2, systems built within an emergent view of emotions have so far stayed away from manipulating emotion in the system, e.g., via models, and instead provide resources — such as visualizations of physiological data (Stahl et al., 2009), or distorted live video streams (Boehner et al., 2008) — for the users to make sense of. This move positions the users as the driver of and solely in charge of emotional interpretation and, implicitly, the system in a passive role.

Freaky walks a delicate line between these two types of systems. In an important sense, it recombines elements from both perspectives. On the one hand, it engages efficient representations of human emotions, i.e., models. On the other hand, it retains the possibility for rich experiences and open-ended user interpretations typical of emergent systems. Put differently, Freaky demonstrates how a balance between human and machine interpretation may be achieved in practice. The system's interaction design allows the user and the system to take turns driving the interaction, sometime seamlessly, other times more abruptly, and keeps the user ultimately in charge of meaning making.

What we seem to have accomplished through Freaky's design is a more nuanced positioning of models of emotion, so that they may participate in the world by provoking and informing users' perspectives rather than replacing them. This was achieved by **contextualizing the models and exposing them**. Thus, the interaction does not revolve around models structuring users' experiences as in the classic, representationalist, view. While the models do shape the interaction in important ways, a crucial aspect of interaction consists of drawing on users' experiences to examine the models, i.e., using familiar experiences as a means to explore, understand, appropriate, and contest physiological, statistical, scientific, or technical approaches to emotion. In other words, interacting with Freaky means blending users' perspectives and formal representations in action. Thus, Freaky demonstrates a different way of designing interaction: one that links human and machines perspectives in and through performances.

Success and Failure

The interaction with Freaky works by exposing and exploring the relationship between different cuts: specifically, between machine and human perspectives on emotion. Thus, ways in which they do or do not line up, elements come together or separate, or boundaries come to be drawn differently are all aspects that define the interaction space afforded by the system. In this space, useful interactions may emerge between Freaky and its users provided that the human and machine perspectives are linked in interaction, and so enable users to engage the machine perspective in (re)figuring their own. A successful interaction, then, equates to performances emerging from linked perspectives so that difference and sameness may be explored and reworked. In short, a system can be said to work from the perspective I take here to the degree in which the user includes the system interaction in the apparatus with which s/he cuts, i.e., sees, the world.

An avenue for future work would involve extending the linking of machine and human perspectives to include the reverse situation: including users' ways of cutting the world in the system's. In the context of machine learning and human emotion this might be achieved by leveraging live user feedback to dynamically update (retrain) models, i.e., on-the-fly learning, already technically feasible.

Having specified what a successful system might mean, I now turn to consider what failure might mean in this framework. Failure would mean that the machine and human perspectives are not linked in interaction. One of Freaky's usage instances point to such a scenario. S used the device while walking to and from a doctor's appointment. A researcher in the hard sciences, she had

mentioned before using Freaky that she had high hopes that she would learn something new about herself. While I was describing how Freaky works and interacts with the user, S looked quite confused and candidly commented: *“so this is a humanistic investigation of fear”* During her time with Freaky, the device detected fear once due to heart rate changes which she saw as arising merely from physical exertion while walking up a steep hill. Otherwise, she confessed to having had very different expectations; she had imagined that the device would give her some *“hard facts sensed or inferred by the systems [laughs] I don’t know like adrenaline levels.”* Her comments suggest that she would have liked to have access to visualizations of the sensed data, to the classification model, to measures of prediction confidence, and so forth. Furthermore, she made it clear that she found the system quite strange and did not relate to it. She thought this might have to do with her training and work as a molecular biologist mentioning, for example, that she was not comfortable with the kind of open-ended interpretations that interacting with Freaky might entail. Because of that she concluded: *“I can’t really imagine a scenario in which the system as it is could have anything interesting to say to me.”* In her case, then, S’s ways of making sense of the world did not incorporate Freaky’s reactions and behaviors.

S’s experience with the system highlights that matters of design play a crucial role in shaping what is made available to users in interaction and how, just as technical and scientific decisions shape representations of emotion and emotion itself (see chapter 4). In her case, the interaction with the system got in the way of S’s making sense of the information available to Freaky (e.g., user heart rate and the model’s prediction). Consequently, a connection with the machine’s perspectives that might have informed S’s views did not materialize. In sum, failure in this framework has to do with human and machine perspectives

not meeting in interaction.

Other Consequences and Exclusions

S's experiences with the system highlight aspects implicitly sidelined by Freaky's design and construction. Specifically, she would have liked the interaction with the device to include having access to the guts of the system, as it were. While the visibility and exposure of the model's predictions made it possible for the majority of the users to engage with the system, the level of detail was still lacking for others. In S's words and her qualifying the system as 'humanistic,' in contrast to her scientific background, we can almost feel her disappointment that *definitive, certain* answers were marginalized by the way we designed the system.

On the topic of visibility some users voiced their concerns that they may not always be comfortable if their reactions are exposed by the system. For example, Uma commented that while she was comfortable with Freaky's reaction in Tyler's presence, she may have felt exposed if other people had been present. These comments point out potential consequences of the model's visibility. This is connected to Freaky's behavior contributing to forming events in the world. In other words, the system brings to light and thus makes real a physiological event that may have gone unnoticed.

Other aspects that the system excluded from materializing or, at a minimum, marginalized were related to the spectrum of emotions towards which Freaky shifted the interaction. As evidenced by user comments, fear and high energy emotions were central, while low emotions were not. This is a consequence of

our decision to orient the model towards intense physiological experiences, as discussed earlier in this chapter.

Wrapping up Evaluation

Here, I return briefly with some meta comments on practices of evaluation. I prefaced the analysis of the observations that emerged from Freaky's user study with a discussion in which I highlighted how an agential realist take on evaluation is different than the classical one(s). The main point stressed there was that the way we do evaluation fundamentally depends on the metaphysical framework adopted (explicitly, in the case of agential realism, or implicitly, in the traditional one). Consequently, the metaphysical perspective shapes the outcomes emerging from evaluation practices, i.e., what we see depends on the lens adopted. In other words, practices of evaluation can be thought to effect a cut in the phenomenon under investigation: the system's interactions with users. The observations and knowledge that emerge from such practices constitute the outcome of the cut. This is to say that the **outcomes of evaluation must be understood as another form of representation**. Therefore, general aspects relevant to representations that surfaced in chapter 4 apply to these descriptions as well:

- We, researchers or practitioners, are part of the phenomenon from which representations emerge, rather than external to it. This means that our metaphysical position must be part and parcel of the analysis. Importantly for matters of evaluation, the outcome depends on the practices that enact our metaphysical perspectives. Thus, an opportunity to better understand how an achievement of evaluation — the results, e.g., a causal relationship

— emerges is to extend our analysis tools to include the interplay between our actions and their underlying philosophical beliefs.

- There isn't one way of performing evaluation, but many, situated, partial perspectives. This means that any evaluation only tells a part of the story: it speaks about one way of reading and doing reality.
- What is left out from these representations is constitutive of reality just as much as what is included. Therefore we must understand also what is left out and examine the consequences of doing so. This insight ties in with the understanding that the outcomes of evaluations are creative of reality: they simultaneously describe and shape reality. **Ethical** concerns, then, must be taken into account: we are responsible for the realities we help enact.

Apart from providing a case study of what evaluation might look like when done within a performative framework, my aim with the user study has been to clarify conceptual matters pertaining to practices of evaluation. In recent years, matters of epistemology have been shown as crucial to questions of evaluation (e.g., Boehner et al., 2007; Kaye, 2009). Through Freaky's user study I have begun to show that matters of ontology, i.e., how we compartmentalize the world, bear important consequences on the specifics of evaluation and its outcomes. Importantly, the fact that different perspectives materialize differently in the world makes us responsible for the differences we help introduce. That we have alternatives to choose from means picking one over another is also a matter of politics and of ethics. In sum, then, my argument in this section is that it is imperative that we extend the scope of analysis to include metaphysical considerations when evaluating systems in use.

5.7 Conclusion

In this chapter, I approach system design, construction, and evaluation within the agential realist framework. As I have described in chapter 3, a fundamental point of departure from the traditional perspective concerns matters of ontology: what we take the world to be and how we compartmentalize it. In contrast to the classic view that sees the world as a collection of entities given in advance, agential realism proposes a relational ontology, one that figures entities and relations as achieved through specific practices than as continuously enacted separations. This difference has important consequences on our understanding of the nature of reality and consequently on the role our practices of knowing and doing should play. In the classical framework, what is real (and therefore true) is given in advance independent of our observations; therefore the aim of investigation is to uncover what has been there all along. Agential realism, on the other hand, understands reality as performed; as such entities and relations may be enacted differently through different practices, including human practices of knowing. Due to these conceptual differences, I approached differently all the steps involved in the construction of the interactive system described in this chapter: design, technical construction (hardware and software), and evaluation.

Building on the research presented in chapter 4, in the technical construction of the system I move away from issues of correspondence between representation and represented that are a dominant concern in the traditional worldview and instead focus on deriving representations that contribute in specific ways to the enactment of emotion. Concretely, instead of aiming to derive representations that capture users' true emotion, I actively crafted representations of

emotion so as to orient the interaction with the system towards physiological events of interest.

The aim of interaction design shifted away from a correct response to the identified user emotion and towards supporting dynamic human-machine performances of emotion. As entities and attributes of interest are in flux, we moved away from static relationships and immutable meanings and towards dynamic co-construction of meanings. We did this by designing the interaction as a ritual through which human and machine perspectives may be connected in interaction so that similarities and differences may be explored and reworked.

In the system evaluation the change in framework brings about a change in the kind of entities traced and analyzed from fixed to dynamic entities and relations to between them. The outcomes of evaluation depend on the conceptual framework adopted in the design and evaluation of the system. Therefore, practices of evaluation must take into account how the conceptual framework contributes to the system's achievements that emerge from the user study. Finally, evaluation must also analyze the kinds of realities that are and are not made possible by the interplay between the specifics of the system and the conceptual framework from which it was built.

CHAPTER 6

CONCLUSION

The structure of this thesis reflects the long and meandering path I have taken to tackle the problems laid out in the introduction chapter. Now, at the end, it is time to reflect where I started, what has happened along the way, and how it all fits together.

I started this dissertation with two issues: the problem of representation and the socio-technical gap. The former pertains to the nature of the relationship between representation and that which is represented. Specifically: do representations simply describe the represented or do they also shape it? The latter speaks of the gap between social requirements for system design and the technical means to support them and describes the limitations associated with this disjuncture. In chapters two and three, we saw that HCI understands the gap as real and inherent in the world.

Representation issues come up in discussions and analyses of the gap. While the representationist approach to representation — which takes representations to simply mirror the represented — currently dominating technical practice has been theorized as problematic, it is not clear what it might mean to represent differently in technical practice. For example, as we have seen in chapter 1, alternative takes on representation raise big questions pertaining to objectivity and truth. Consequently, it is not clear whether a rigorous and systematic alternative technical practice is possible. At the beginning of this thesis, then, we were stuck with a gap that is inherent in the world and a representationist take on representation. Further, the nature of the relationship between the representation problem and the gap was unclear.

This dissertation shows the two issues — the socio-technical gap and the representation problem — go hand in hand. Moreover, they stem from and are thus entangled with the web of metaphysical assumptions that underlie most research in HCI. This understanding has been a major milestone in my work because it generates fundamentally different insights into these issues.

The relationship between representation and represented depends fundamentally on the metaphysical framework one inevitably relies on, implicitly or explicitly. Of particular importance are the assumptions about the relationship between knower and world implied by these frameworks. In the traditional metaphysical belief system which invisibly underlies most of the work in HCI and computer science, more broadly, the knower is considered to be *a priori* external to the object of study. The very distance between subject and object is the basis for objectivity in this framework. The distance makes possible a similar distance between representation and represented, i.e., between knowledge and the world. Because the subject (e.g., researcher) is located outside the object of study, the representation/knowledge derived is similarly located outside the object: it does not change the object and it aims to mirror it as much as possible. In other words, the knowledge obtained occupies a separate and distinct ontological realm from that of the object or phenomenon it describes: knowledge of the world and the world are separate and distinct categories. One of this framework's implications for technical work is that technical representations should mirror their correspondents and that deriving these representations, if done correctly, do not change the represented in constitutive ways.

However, alternative metaphysical belief systems do exist. The one I have drawn upon in this dissertation — Barad's agential realism (Barad, 2007) — un-

derstands the subject to be part of the world, rather than inherently external to it. In Barad's framework, specific practices are needed to achieve the separation between subject and object. Furthermore, different actions may lead to different separations. Because different separations between object and subject are possible, objectivity cannot simply mean that the subject be external to the object of study. Rather, objectivity must be rethought to include the ways in which the actions performed achieved the specific separation observed. In this alternative framework, there is no inherent distance between subject and object and similarly between representation and represented. Rather, subject and object, representation and represented are related in constitutive ways and the exact relationships enacted between them depend on the very actions that enacted them.

Contrasting the views on representation emerging from the two frameworks highlights important differences. The assumptions within the traditional world-view figure representation at a distance from the represented, meaning that the separation between the two is *inherent* in the world. Therefore, **performing representations does not determine the represented and the former should mirror the latter**. The alternative framework, agential realism, sees the subject as always already entangled in the world. In order to construct representations, then, **separations between subject and object must occur**. The actions that achieve the separations contribute towards *forming* the object of study. Importantly, then, the interplay between subject and object is constitutive of the two entities. It follows that the interplay between knowledge and world (in our case, representation and represented) is also constitutive.

These are two takes on representation: one for each framework. The representationist take is made possible by the traditional framework. The alternative framework, agential realism, sees representation as simultaneously shaping and describing the represented — just as in the physiological maps project I used as an example at the start of this dissertation. It now becomes clear that taking a different approach to representation invokes a different set of metaphysical assumptions. We now understand that the questions on objectivity and truth raised in response to my work stemmed from the clash between the metaphysical assumptions underlying the different take on representation I was experimenting with and those in the classical framework. Making sense of them together was impossible. Seeing this required making visible *both* the traditional framework and finding a new one to serve as an alternative that would understand the nature of representation differently.

Now we have an alternative understanding of representation together with a framework which renders it coherent and allows for systematic work to be done. One question remains: **what is the nature of the socio-technical gap in this alternative framework?**

I answer this question by grounding the discussion in the case study presented in chapter 5 detailing Freaky's design and user interactions with it. My strategy here is to contrast the views that emerge from the two frameworks so that we may glean the novel possibilities afforded by the agential realist framework. When we work within the classic framework, we understand users as inherently separate from Freaky. Users' and Freaky's views on emotion are separate from another, i.e., the user experiences an emotion which the system attempts to decode from heart rate data. In other words, the boundaries be-

tween the two perspectives are clear. Comparing their respective perspectives¹ on emotion reveals a gap between them, i.e., between experienced emotions and Freaky's take on them via statistical classification of heart rate information into fear or not fear.

But users' interactions with Freaky, which I presented in chapter 5, show that disentangling users' emotions from Freaky's actions may not always be possible. Recall that Freaky's mere presence has users accounting differently for their experiences. This suggests an entanglement between users and Freaky, in other words, between elements of the social and the technical. The performative framework makes space for this entanglement to be examined: users' and Freaky need not be separate in every aspect; in fact, this entanglement points to elements of wholeness: instances in which user and Freaky may be thought of as partially and temporarily fused together. The performative framework, then, allows us to talk about Freaky and users both as fused and as separate entities. Thus it accounts for phenomena that the classical framework cannot account for and therefore become invisible or recurrent problems, i.e., perpetually figured as something to be dealt with in future work.

While figuring users and Freaky as partially and temporarily fused, the agential realist framework also makes possible talking about important separations. These have to do with the differences between Freaky's and users' perspectives on emotion. For example, users' perspectives may be informed by their emotional past to which Freaky may not have access to; while Freaky's rests on an ability to make statistical comparisons between physiological data, which users cannot do. These differences might be considered socio-technical

¹The classical framework allows for two possible explanations for such differences: either Freaky's recognition of user's emotion is wrong, or Freaky correctly recognized fear but the user is not aware that s/he is scared or does not want to admit it.

gaps.

As highlighted by the performative analysis of users' interactions with Freaky, these gaps are in flux as the relationship between users' and Freaky's perspectives is dynamic. Fusions and separations between users' and Freaky's points of view are continuously enacted. This enables the reworking in interaction of differences and similarities between their respective views of emotion. In the performative framework, we observe many temporary gaps, continuously in flux. Crucially, these temporary, dynamic gaps act as a magnet pulling users to examine and rework the contours of their perspective in dialogue with the machine's. This dynamic suggests design strategies to engage the gaps in productive ways: exposing them so that they may be examined, engaged, and transformed. In sum, in the performative framework the gaps emerge as generative and resourceful, rather than a pesky feature of the world whose effects need to be constantly minimized.

In conclusion, perhaps the most important point here is that how we understand and approach the gap depends fundamentally on the metaphysical framework we adopt, implicitly or explicitly. If we position ourselves within the traditional worldview we understand separations as inherent in the world. Therefore, the separation between the social and technical (like that between subject and object) appears inherent in the world and consequently the distance between the two (the gap) must also be located in the world (i.e., independent of our actions). Furthermore, we will see as legitimate only those actions (e.g., research, designing, building, evaluating) that appear to respect the allegedly inherent separation and fall neatly within one camp (social) or the other (technical). In other words, we shall stay clear of actions that blur the boundary

because they are illegal operations in this framework, i.e., they go against the most fundamental assumptions that make up the framework.

In contrast, the performative framework sees separations as performed rather than inherent in the world. Different actions achieve different separations, i.e., we can separate the social and the technical in different ways. The gap(s) between the two depend on the concrete actions that continuously enact the separations. Hence the gaps are always in flux: while some may disappear, others necessarily appear. In this framework, the gaps are the effect of separations. Seeing them as such suggests performing different separations so that they may be transformed. Crucially, then, gaps appear as generative within the agential realist framework.

Summary of the key points:

- The two issues I started with (the problem of representation and the socio-technical gap) are entangled within an implicit metaphysical framework. To approach these issues differently, we need a different framework. I draw on one such alternative framework, Barad's agential realism.
- The two issues flip together: Changing our approach to one of the issues (e.g., letting go of representationalism) must be accompanied by a change in framework, which necessarily entails a different take on the other issue (e.g., the gap).
- In the performative framework:
 - Representations shape and describe
 - Gap depends on specific separations; it is always in flux and generative.

6.1 Limitations

In this dissertation, I have shown that taking a fundamentally different metaphysical position can be fruitful for HCI practice. Necessarily, this research has a number of limitations. Here, I take a moment to reflect on them and on how I might have done things differently. First, as the research makes clear, adopting a different metaphysical position requires a major overhaul at the conceptual level and reworking *every* aspect of HCI within this new position. Such reworkings happen gradually. Consequently, in some stages of the work presented here it was hard to grasp what in the new framework is important and how HCI practices should morph. The work presented here and the writing reflect these challenges. For example, designing and building Freaky went on in parallel with figuring things out conceptually. In its current version, the part in chapter 5 on Freaky's technical construction does not do justice to the sinuous, trial and error based process from which Freaky emerged. As a result, intricate performances (including mine) disappear from view. These are important, indeed constitutive, in the performative framework I adopted here. In the classical framework, however, focusing primarily on what worked (the end result) to the detriment of all the other less fruitful paths is the *modus operandi* (recall from chapter 4 that one of the consequences of the classical framework for technical practice is that the craft work integral to this kind of work disappears from view). Writing myself out of parts of the thesis attests just how difficult it is to leave behind the traditional metaphysical framework. Were I to go through the process again I would keep track on my positions throughout, the choices I made, the exclusions I enacted. Put it differently, I would include myself more in the accounts presented here.

Undoubtedly, there are other remains of the old framework in this dissertation. I trust they will surface in time, in my or others' future work.

6.2 A Critical Technical Practice

In chapter 4 I began telling you that this project, working out the implications of Barad's agential realism for HCI, is a critical technical practice. I conclude this dissertation by succinctly considering how the work presented here builds on and extends Agre's critical technical practice (CTP).

Agre argues for the necessity of CTP so that "rigorous reflection upon technical ideas becomes an integral part of day-to-day technical work itself" (Agre, 1997). He posits that every technical field is founded upon a base of commonly accepted theories or systems of thought. These theories embody particular cultural assumptions, which usually go undetected by the field's practitioners. By uncovering and altering those assumptions, technical impasses can be accounted for and perhaps solved. Barad's framework resonates with this in that it reworks foundational assumptions, such as metaphysical ones. In this thesis, I have traced the implications of such assumptions in HCI practice (focusing specifically on the socio technical gap and the problem of representation) and worked out how we might do HCI differently by working out on the ground the implications of a new set of assumptions.

While metaphysical matters (e.g., ontology) clearly inform Agre's work (i.e., in his writings he implicitly positions his work within what appears to be a relational metaphysics), he does not tackle such matters explicitly. From this perspective, Barad's framework makes possible extending what we mean by

and how we do CTP.

Importantly, one such extension has to do with the space of possibilities, moving forward, reconfiguring the world, including getting unstuck, i.e., overcoming recurrent difficulties. To make this difference clear, a bit more detail on Agre's work is needed. In his model of technical practice, Agre understands the theories and practices of a field to be structured by generative metaphors: metaphors which allow the field to extend its discourse by explaining its practices. Generative metaphors rely on a hierarchical opposition: a classification of elements into two categories: one that is central and is considered well-behaved (e.g., supporting a theory) and the other peripheral and ill-behaved (not conforming to the considered theory). Agre calls these categories center and margin. Theories often highlight the centers and hide the margins in footnotes, exceptions, or assumptions. To move a technical field forward, Agre's model produces a new generative metaphor by reversing the roles of center and margin.

While offering first steps towards repositioning technical practices, Agre's model may appear to view the world in a static way. What happens when the new metaphor runs its course? Would a further reversal bring back the old metaphor? A technical practice founded on Barad's agential realism approaches the space of possibilities as open ended and continuously in flux, thus accounting for the ongoing dynamism of the world. This dynamism is fundamental to her metaphysics, one that sees the world continuously and iteratively remade through practices of engagement. Crucially, agential realism puts forward a dynamic way of approaching the space of possibilities: "possibilities do not sit still. [] Possibilities aren't narrowed in their realization; new possibilities open

up as others that might have been possible are now excluded: possibilities are reconfigured and reconfiguring. There is a vitality to intra-activity, a liveliness, not in the sense of a new form of vitalism, but rather in terms of a new sense of aliveness. The world's effervescence, its exuberant creativeness can never be contained or suspended. Agency never ends; it can never "run out."" (Barad, 2007, p. 234-5)

6.3 Looking Ahead

In my future work, I intend to further develop an agential realist position in HCI. This will necessarily entail understanding further how important separations and fusions may be achieved, including but not limited to the ones between users and systems. Related to this point, I wish to return to the central concept of agency and working out in detail through technical and design case studies what it might mean for HCI to approach agency as an inexhaustible force (not unlike gravity), rather than an attribute inherent in entities. Finally, I would like to continue investigating how to include ourselves, the researchers, and our situated (e.g., analytical) positions more in the knowledge we help articulate.

In line with this last point, perhaps the most important contribution of my PhD research is to offer a detailed explanation and demonstrations of how, in HCI, our implicit views of the world guide in important ways our actions and, in turn, these help shape the world itself. This generates a dynamic through which what we put out there comes back to us, as it were. This insight is closely linked to another: it is possible to see and thus do the world differently, al-

beit difficult at first. For me, the research presented in this dissertation brought about a new sense of freedom and, at the same time, responsibility. I came to understand that, on the one hand, we are no longer bounded by constraints we once believed to be immutable and thus we may create alternatives outside of them; on the other, we are now even more accountable for the knowledge, the systems, the experiences, and the user practices we help create. I end hoping the accounts I presented here will be useful to the HCI community and will further enable it to participate in the world in creative, responsive, and responsible ways.

APPENDIX A

SVM PARAMETER TESTING

The parameters of SVM^{light} tested included $-c$, $-j$, t and T : c is the trade-off between training error and the size of margin; j is the cost factor by which training errors on positive examples outweigh errors on negative examples; t is the type of kernel function used; and, T is the window size.

A.1 Kernel Type

The predictions from the 40 models (10 bagged train sets times four participants) were averaged. The results are summarized in the following table. The baseline accuracy was 68.5%.

kernel type	Accuracy	Precision	Recall
linear	82.3%	74.8%	65.7%
polynomial	88.4%	82.4%	79.4%
radial basis	89.2%	82.2%	82.8%
sigmoid	63.0%	41.5%	39.8%

Table A.1: Summary of results for kernel type parameter testing.

The radial basis kernel turned out the highest percentages. The polynomial kernel had slightly higher precision rate, but the radial basis kernel outperformed the polynomial kernel in both accuracy and recall.

A.2 Time Window Size

We trained using the radial basis kernel. The predictions from 10 models from the 10 bagged train sets were averaged. The baseline accuracy was 77.2%. The results are summarized below:

T	Accuracy	Precision	Recall
20	88.9%	72.3%	83.3%
100	89.7%	75.6%	81.2%
200	90.4%	78.1%	80.7%
1000	88.2%	77.7%	68.4%

Table A.2: Summary of results for window size parameter testing.

The window size of 200 worked the best; it had the best accuracy and precision rate. The recall rate dropped as time window size increased.

A.3 Trade-Off C

We used $T = 200$ and trained using the radial basis kernel. Due to time constraints, only three models were trained for each participant, so the result is an average of 12 different models in total. The baseline accuracy was 68.5%.

The C value of 0.5 produced the best accuracy and precision rate, but the recall rate was too low. The recall rate increased as C value increased, until it got as large as 1000. Ignoring $C = 0.5$, the $C = 500$ resulted in the best accuracy, precision, and recall rates. Thus, we used $C = 500$ for the following experiments.

C	Accuracy	Precision	Recall
0.5	93.1%	91.3%	77.5%
1	88.5%	82.0%	80.5%
5	90.6%	84.8%	84.0%
10	91.3%	85.7%	85.5%
20	91.6%	85.7%	85.5%
30	91.9%	86.8%	86.4%
40	92.1%	87.2%	86.9%
100	92.4%	87.8%	87.5%
500	92.5%	87.9%	87.6%
1000	92.3%	87.5%	87.2%

Table A.3: Summary of results for trade-off parameter testing.

A.4 Cost Factor J

The training sets used for this parameter testing were the same as in step 3, i.e. average of 12 different models in total, with $T = 100$ trained using the radial basis kernel. The trade off parameter c was set to 500. The baseline accuracy was 68.5%.

J	Accuracy	Precision	Recall
0.5	92.2%	88.6%	85.5%
1	92.5%	87.9%	87.6%
1.5	92.4%	87.3%	87.9%
2	92.3%	87.0%	88.3%

Table A.4: Summary of results for cost factor parameter testing.

As j increased, there was a trade-off between precision and recall. The accuracy was maximized at $j = 1$ (the default value for j). Therefore, we chose $j = 1$ as the best cost factor value.

BIBLIOGRAPHY

- Aarts, E. (2004). Ambient intelligence. In P. De Bra & W. Nejdl (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems*, volume 3137 of *Lecture Notes in Computer Science* (pp. 548–568). Springer Berlin / Heidelberg.
- Abowd, G. D. & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Interact.*, 7, 29–58.
- Ackerman, M. S. (2000). The intellectual challenge of cscw: the gap between social requirements and technical feasibility. *Hum.-Comput. Interact.*, 15, 179–203.
- Agre, P. E. (1997). *Computation and Human Experience*. Cambridge University Press.
- Ark, W. S. & Selker, T. (1999). A look at human interaction with pervasive computers. *IBM Systems Journal*, 38, 504–507.
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions.
- Bannon, L. J. (1995). The politics of design: Representing work. *Commun. ACM*, 38(9), 66–68.
- Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs*, 28(3), 801–831.
- Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.

- Bardram, J. E., Frost, M., Szántó, K., & Marcu, G. (2012). The monarca self-assessment system: a persuasive personal monitoring system for bipolar patients. In *IHI* (pp. 21–30).
- Bassoli, A., Brewer, J., Martin, K., Dourish, P., & Mainwaring, S. D. (2007). Underground aesthetics: Rethinking urban computing. *IEEE Pervasive Computing*, 6(3), 39–45.
- Blanchette, J. (2010). A Material History of Bits. *polaris.gseis.ucla.edu*, (pp. 1–30).
- Boehner, K., de Paula, R., Paul, & Sengers, P. (2007). How emotion is made and measured. *Int. J. Hum.-Comput. Stud.*, 65(4), 275–291.
- Boehner, K. & Sengers, P. (2006). Closing the Affective Gap & Opening Up Evaluation. In *HUMAINE Workshop on Innovative Approaches for Evaluating Affective Systems*.
- Boehner, K., Sengers, P., & Warner, S. (2008). Interfaces with the ineffable: Meeting aesthetic experience on its own terms. *ACM Trans. Comput.-Hum. Interact.*, 15, 12:1–12:29.
- Bowker, G. C. & Star, S. L. (1999). *Sorting things out : classification and its consequences*. Cambridge, Mass.: MIT Press.
- Brooks, R. A. (1991). Intelligence without representation. *Artif. Intell.*, 47(1-3), 139–159.
- Butler, J. (1993). *Bodies That Matter: On the Discursive Limits of Sex*. Routledge.
- Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *ICML* (pp. 161–168).

- Chalmers, M. (2004). A historical view of context. *Computer Supported Cooperative Work*, 13(3), 223–247.
- Chang, C.-C. & Lin, C.-J. (2001). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1–27:27.
- Coan, J. A. & Allen, J. B. e. (2007). Introduction: Organizing the tools and methods of affective science. *The handbook of emotion elicitation and assessment*, (pp. 3–7).
- Collet, C., Vernet-Maury, E., Delhomme, G., & Dittmar, A. (1997). Autonomic nervous system response patterns specificity to basic emotions. *J. Auton. Nerv. Syst*, 62(1-2), 45–57.
- Crosby, M. E. & Auernheimer, B. (2001). Physiological data feedback for application in distance education. In *In Proceedings of PUI 2001*. NY: ACM. Pp (pp. 1–5).
- Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Kalin, N. H. (2003). Parsing the subcomponents of emotion and disorders of emotion: perspectives from affective neuro-science. *Handbook of Affective Sciences*, Davidson, R. J., Scherer, K. R., Goldsmith, H. H. (eds.).
- Dey, A. K. (2001). Understanding and using context. *Personal Ubiquitous Comput.*, 5, 4–7.
- Dey, A. K., Abowd, G. D., & Salber, D. (1999). A context-based infrastructure for smart environments.
- Dourish, P. (2001). *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press.

- Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1), 19–30.
- Dourish, P. (2006). Implications for design. In *CHI* (pp. 541–550).
- Ferreira, P., Sanches, P., Höök, K., & Jaensson, T. (2008). License to chill!: how to empower users to cope with stress. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges, NordiCHI '08* (pp. 123–132). New York, NY, USA: ACM.
- Galison, P. (1995). Objectivity is romantic. *ACLS Occasional Paper*, 47, 15–43.
- Gaver, W. W., Sengers, P., Kerridge, T., Kaye, J., & Bowers, J. (2007). Enhancing ubiquitous computing with user interpretation: field testing the home health horoscope. In *CHI* (pp. 537–546).
- Gray, E. K. & Watson, D. (2007). Assessing positive and negative affect via self-report. *The handbook of emotion elicitation and assessment*, (pp. 171–183).
- Gross, J. J. & Levenson, R. W. (1995). Emotion elicitation using films. *Cog. and Emot.*, 9(1), 87–108.
- Haraway, D. (1988). SITUATED KNOWLEDGES : THE SCIENCE QUESTION IN FEMINISM OF PARTIAL AND THE PRIVILEGE. *Feminist Studies*, 14(3), 575–599.
- Healey, J., Nachman, L., Subramanian, S., Shahabdeen, J., & Morris, M. E. (2010). Out of the lab and into the fray: Towards modeling emotion in everyday life. In *Pervasive* (pp. 156–173).
- Healey, J. & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. volume 6 (pp. 156–166).

- Höök, K. (2008). Affective loop experiences - what are they? In *PERSUASIVE* (pp. 1–12).
- Höök, K., Stahl, A., Sundström, P., & Laaksolahti, J. (2008). Interactional empowerment. In *CHI* (pp. 647–656).
- Intille, S. S., Bao, L., Tapia, E. M., & Rondoni, J. (2004). Acquiring in situ training data for context-aware ubiquitous computing applications. In *CHI* (pp. 1–8).
- Jacovi, M., Soroka, V., Gilboa-Freedman, G., Ur, S., Shahar, E., & Marmasse, N. (2006). The chasms of cscw: a citation graph analysis of the cscw conference. In *CSCW* (pp. 289–298).
- Joachims, T. (1998). Making large-scale support vector machine learning practical.
- Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Man-Machine Studies*, 65(8), 724–736.
- Kaye, J. (2009). *The Epistemology and Evaluation of Experienced-focused HCI*. PhD thesis, Cornell University, Ithaca, NY, USA.
- Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., MacIntyre, B., Mynatt, E. D., Starner, T., & Newstetter, W. (1999). The aware home: A living laboratory for ubiquitous computing research. In *CoBuild* (pp. 191–198).
- Kientz, J. A., Boring, S., Abowd, G. D., & Hayes, G. R. (2005). Abaris: Evaluating automated capture applied to structured autism interventions. *ubicomp*.
- Klemmer, S. R., Hartmann, B., & Takayama, L. (2006). How bodies matter: five themes for interaction design. In *Conference on Designing Interactive Systems* (pp. 140–149).

- Laird, J. D. & Strout, S. (2007). Emotional behaviour as emotional stimuli. *The handbook of emotion elicitation and assessment*, (pp. 54–63).
- Latour, B. (1993). *We Have Never Been Modern*. Harvard University Press.
- Leahu, L., Schwenk, S., & Sengers, P. (2008a). Subjective objectivity: Negotiating emotional meaning. In *Conference on Designing Interactive Systems (DIS'08)* (pp. 425–434).
- Leahu, L., Sengers, P., & Mateas, M. (2008b). Interactionist AI and the promise of ubicomp, or, how to put your box in the world without putting the world in your box. In *UbiComp'08* (pp. 134–143).
- Li, Y. & Landay, J. A. (2008). Activity-based prototyping of ubicomp applications for long-lived, everyday human activities. In *CHI* (pp. 1303–1312).
- Lisetti, C. L., Nasoz, F., LeRouge, C., Ozyer, O., & Alvarez, K. (2003). Developing multimodal intelligent affective interfaces for tele-home health care. *Int. J. Hum.-Comput. Stud.*, 59(1-2), 245–255.
- Lo, J., yun Lin, T., hua Chu, H., chin Chou, H., hao Chen, J., jen Hsu, J. Y., & Huang, P. (2007). Playful tray: Adopting ubicomp and persuasive techniques into play-based occupational therapy for reducing poor eating behavior in young children. In *in Proceedings of the 9th International Conference on Ubiquitous Computing (UBICOMP)*.
- Mandryk, R. L., Atkins, M. S., & Inkpen, K. M. (2006). A continuous and objective evaluation of emotional experience with interactive play environments. In *CHI* (pp. 1027–1036).
- Nold, C. (2004). Bio mapping/emotion mapping.
<http://http://biomapping.net/>.

- Park, S. H., Won, S. H., Lee, J. B., & Kim, S. W. (2003). Smart home digitally engineered domestic life. *Personal and Ubiquitous Computing*, 7, 189–196. 10.1007/s00779-003-0228-9.
- Picard, R. W. (1997). *Affective Computing*. MIT Press.
- Picard, R. W. & Healey, J. (1997). Affective wearables. *Personal and Ubiquitous Computing*, 1(3), 231–240.
- Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(10), 1175–1191.
- Pousman, Z., Romero, M., Smith, A. M., & Mateas, M. (2008). Living with tableau machine: a longitudinal investigation of a curious domestic intelligence. In *UbiComp* (pp. 370–379).
- Ramirez, L. & Dyrks, T. (2010). Designing for high expectations: balancing ambiguity and thorough specification in the design of a wayfinding tool for firefighters. In *Conference on Designing Interactive Systems* (pp. 390–399).
- Rode, J. A. (2011). Reflexivity in digital anthropology. In *CHI* (pp. 123–132).
- Rogers, Y. (2006). Moving on from weiser’s vision of calm computing: Engaging ubicomp experiences. In *Ubicomp* (pp. 404–421).
- Romero, M., Pousman, Z., & Mateas, M. (2008). Alien presence in the home: the design of tableau machine. *Personal and Ubiquitous Computing*, 12(5), 373–382.
- Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.

- Salber, D., Dey, A. K., & Abowd, G. D. (1999). The context toolkit: aiding the development of context-enabled applications. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, CHI '99* (pp. 434–441). New York, NY, USA: ACM.
- Sengers, P. (2006). Closing the Affective Gap. In *HCII Seminar Presentation at Carnegie Mellon University*.
- Sengers, P., Boehner, K., Mateas, M., & Gay, G. (2008). The disenchantment of affect. *Personal and Ubiquitous Computing*, 12(5), 347–358.
- Sengers, P. & Gaver, B. (2006). Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Conference on Designing Interactive Systems* (pp. 99–108).
- Sinha, R. & Parsons, O. (1996). Multivariate response patterning of fear and anger. *Cog. and Emot.*, 10(2), 173–198.
- Stahl, A., Höök, K., Svensson, M., Taylor, A. S., & Combetto, M. (2009). Experiencing the affective diary. *Personal and Ubiquitous Computing*, 13(5), 365–378.
- Suchman, L. (1987). *Plans and Situated Actions: The problem of human-machine communication*. Cambridge University Press.
- Suchman, L. (2006). *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press.
- Tapia, E. M., Intille, S. S., & Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive* (pp. 158–175).
- Tennenhouse, D. (2000). Proactive computing. *Commun. ACM*, 43, 43–50.
- Verran, H. (2001). *Science and an African Logic*. University of Chicago Press.

- Weiser, M. (1991). The computer for the 21st century. *Sci. Am.*, 265, 94–104.
- Wiens, S. & Ohman, A. (2007). Probing unconscious emotional processes. *The handbook of emotion elicitation and assessment*, (pp. 65–83).
- Woolf, B. P., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. G., & Picard, R. W. (2009). Affect-aware tutors: recognising and responding to student affect. *IJLT*, 4(3/4), 129–164.
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '07* (pp. 493–502). New York, NY, USA: ACM.